

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CẦN THƠ  
TRƯỜNG CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG



NIÊN LUẬN NGÀNH KHOA HỌC MÁY TÍNH

Đề tài

XÂY DỰNG HỆ THỐNG XÁC ĐỊNH  
MÓN ĂN ĐẶC SẢN VIỆT NAM  
DỰA TRÊN CÁC THÀNH PHẦN NGUYÊN LIỆU

Sinh viên thực hiện:

Đỗ Hiếu Nghĩa

Mã số: B2016985

Khóa: 46

Cần Thơ, 05/2024

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CẦN THƠ  
TRƯỜNG CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG



NIÊN LUẬN NGÀNH KHOA HỌC MÁY TÍNH

Đề tài

XÂY DỰNG HỆ THỐNG XÁC ĐỊNH  
MÓN ĂN ĐẶC SẢN VIỆT NAM  
DỰA TRÊN CÁC THÀNH PHẦN NGUYÊN LIỆU

Giảng viên hướng dẫn:  
TS. Mã Trường Thành

Sinh viên thực hiện:  
Đỗ Hiếu Nghĩa  
Mã số: B2016985  
Khóa: 46

Cần Thơ, 05/2024

# NHẬN XÉT CỦA GIẢNG VIÊN

Cần Thơ, ngày tháng năm 2024

## LỜI CẢM ƠN

Để hoàn thành bài niêm luận này, tôi xin được bày tỏ lòng biết ơn chân thành và sâu sắc đến Thầy Mã Trường Thành, người đã trực tiếp tận tình hướng dẫn, giúp đỡ tôi. Trong suốt quá trình thực hiện đề tài, nhờ những sự chỉ bảo, động viên và hướng dẫn quý giá từ Thầy mà bài niêm luận này được hoàn thành một cách tốt nhất.

Đồng thời, tôi cũng xin gửi lời cảm ơn chân thành đến các Thầy, Cô trường Đại học Cần Thơ, đặc biệt là các Thầy, Cô ở Trường CNTT&TT, những người đã truyền đạt những kiến thức, giúp tôi tích lũy kinh nghiệm quý báu qua hơn bốn năm đại học để đủ khả năng, kinh nghiệm hoàn thành đề tài: “Xây dựng hệ thống xác định món ăn đặc sản Việt Nam dựa trên các thành phần nguyên liệu”.

Cuối cùng, tôi xin chân thành cảm ơn gia đình, bạn bè đã luôn động viên, khích lệ và tạo điều kiện giúp đỡ trong suốt quá trình thực hiện để tôi có thể hoàn thành bài niêm luận một cách tốt nhất.

Cần Thơ, ngày tháng năm 2024

Người viết

Đỗ Hiếu Nghĩa

## LỜI CAM ĐOAN

Tôi tên Đỗ Hiếu Nghĩa, sinh viên ngành Khoa học máy tính, khóa 46. Tôi xin cam đoan niêm luân ngành với đề tài “Xây dựng hệ thống xác định món ăn đặc sản Việt Nam dựa trên các thành phần nguyên liệu” (Building a system for identifying Vietnamese specialty dishes based on their ingredients) là công trình nghiên cứu của bản thân tôi, được sự hướng dẫn bởi TS. Mã Trường Thành.

Các thông tin được sử dụng tham khảo trong niêm luân được thu thập từ các nguồn đáng tin cậy, đã được kiểm chứng, được công bố rộng rãi và được tôi trích dẫn nguồn gốc rõ ràng ở phần tài liệu tham khảo. Các kết quả nghiên cứu được trình bày trong niêm luân này là do chính tôi thực hiện một cách nghiêm túc, trung thực và không trùng lặp với các đề tài khác đã được công bố trước đây.

Cần Thơ, ngày tháng năm 2024

Người cam đoan

Đỗ Hiếu Nghĩa

## MỤC LỤC

|  |           |
|--|-----------|
| <b>MỤC LỤC .....</b>   | <b>1</b>  |
| <b>DANH MỤC HÌNH .....</b>   | <b>4</b>  |
| <b>DANH MỤC BẢNG .....</b>   | <b>6</b>  |
| <b>DANH MỤC CÁC TỪ VIẾT TẮT .....</b>  | <b>7</b>  |
| <b>TÓM TẮT .....</b>   | <b>8</b>  |
| <b>ABSTRACT .....</b>  | <b>9</b>  |
| <b>PHẦN GIỚI THIỆU.....</b>  | <b>10</b> |
| <b>1. Đặt vấn đề.....</b>  | <b>10</b> |
| <b>2. Lịch sử giải quyết vấn đề .....</b>  | <b>11</b> |
| 2.1. A Real-time Junk Food Recognition System based on Machine Learning (2021)<br>.....                            | 11        |
| 2.2. Xác định món ăn đặc sản Việt Nam với sự kết hợp của mạng học sâu và bản thể<br>học (2023) .....               | 12        |
| 2.3. A Review of Image-based Food Recognition and Volume Estimation Artificial<br>Intelligence Systems (2023)..... | 14        |
| <b>3. Mục tiêu đề tài.....</b>   | <b>15</b> |
| 3.1. Mục tiêu chung.....   | 15        |
| 3.2. Mục tiêu chi tiết .....   | 16        |
| <b>4. Phạm vi nghiên cứu .....</b>   | <b>16</b> |
| 4.1. Đối tượng nghiên cứu .....  | 16        |
| 4.2. Phạm vi nghiên cứu .....  | 17        |
| <b>5. Phương pháp nghiên cứu .....</b>   | <b>17</b> |
| 5.1. Phương pháp nghiên cứu tài liệu.....  | 17        |
| 5.2. Phương pháp nghiên cứu thực nghiệm .....  | 17        |
| <b>6. Kết quả đạt được .....</b>   | <b>18</b> |
| <b>7. Bố cục bài báo cáo .....</b>   | <b>18</b> |
| <b>PHẦN NỘI DUNG .....</b>   | <b>19</b> |
| <b>CHƯƠNG 1 .....</b>  | <b>19</b> |
| <b>MÔ TẢ BÀI TOÁN .....</b>  | <b>19</b> |
| <b>1.1. Mô tả chi tiết bài toán .....</b>  | <b>19</b> |

|   |           |
|---|-----------|
| <b>1.2. Vấn đề và giải pháp liên quan đến bài toán .....</b>                  | <b>20</b> |
| 1.2.1. Các nguyên liệu có trong món ăn đặc sản Việt Nam .....                 | 20        |
| 1.2.2. Object detection .....   | 23        |
| 1.2.3. YOLO.....  | 25        |
| 1.2.4. Yolov8.....  | 26        |
| 1.2.5. Yolov9.....  | 28        |
| 1.2.6. RT-DETR .....  | 30        |
| <b>CHƯƠNG 2 .....</b>   | <b>33</b> |
| <b>THIẾT KẾ VÀ CÀI ĐẶT .....</b>  | <b>33</b> |
| <b>2.1. Thiết kế hệ thống .....</b>   | <b>33</b> |
| <b>2.2. Tập dữ liệu huấn luyện mô hình .....</b>                              | <b>34</b> |
| <b>2.3. Xây dựng mô hình huấn luyện nhận diện thành phần nguyên liệu.....</b> | <b>36</b> |
| <b>2.4. Xây dựng mô hình huấn luyện phân lớp món ăn đặc sản .....</b>         | <b>38</b> |
| <b>2.5. Xây dựng công thức tính khoảng cách .....</b>                         | <b>39</b> |
| <b>2.6. Cài đặt hệ thống.....</b>   | <b>41</b> |
| 2.6.1. Thu thập dữ liệu .....   | 42        |
| 2.6.2. Tiền xử lý dữ liệu .....   | 42        |
| 2.6.3. Phân chia tập dữ liệu .....  | 42        |
| 2.6.4. Huấn luyện mô hình .....   | 43        |
| 2.6.4.1. Yolov8 .....   | 43        |
| 2.6.4.2. RT-DETR.....   | 44        |
| 2.6.4.3. Yolov9 .....   | 46        |
| 2.6.4.4. Các mô hình phân lớp.....  | 48        |
| 2.6.5. Cài đặt giải thuật tính khoảng cách .....                              | 53        |
| 2.6.6. Triển khai mô hình .....   | 54        |
| <b>CHƯƠNG 3 .....</b>   | <b>56</b> |
| <b>KẾT QUẢ, ĐÁNH GIÁ VÀ GIAO DIỆN .....</b>                                   | <b>56</b> |
| <b>3.1. Môi trường thực nghiệm.....</b>                                       | <b>56</b> |
| 3.1.1. Cấu hình máy .....   | 56        |
| 3.1.2. Các thư viện sử dụng.....  | 56        |

|  |           |
|--|-----------|
| <b>3.2. Kết quả kiểm tra, đánh giá .....</b>                           | <b>57</b> |
| 3.2.1. Kết quả kiểm thử mô hình phát hiện thành phần nguyên liệu ..... | 57        |
| 3.2.2. Kết quả kiểm thử mô hình phân lớp món ăn đặc sản.....           | 59        |
| <b>3.3. Giao diện.....</b>   | <b>60</b> |
| <b>PHẦN KẾT LUẬN .....</b>   | <b>62</b> |
| 1. Kết quả đạt được .....  | 62        |
| 2. Hướng phát triển .....  | 62        |

## DANH MỤC HÌNH

|  |    |
|--|----|
| Hình 1: Kiến trúc DarkNet53 của mô hình Yolov3 .....   | 11 |
| Hình 2: Mô hình cho bảo tồn món ăn đặc trưng Việt Nam .....  | 13 |
| Hình 3: Mô tả quy trình hoạt động của hệ thống đánh giá chế độ dinh dưỡng .....  | 14 |
| Hình 4: Quy trình xác định món ăn của hệ thống .....   | 19 |
| Hình 5: Món bún cá ở một số tỉnh thành của Việt Nam.....   | 20 |
| Hình 6: Một số món ăn đặc sản Việt Nam .....   | 21 |
| Hình 7: Một số kết quả nhận diện nguyên liệu của mô hình Yolov8.....   | 24 |
| Hình 8: Phân loại mô hình object detection .....   | 25 |
| Hình 9: Cấu trúc CSPDarkNet53 .....  | 27 |
| Hình 10: PGI và các kiến trúc mạng liên quan.....  | 29 |
| Hình 11: Kiến trúc tổng thể của mạng GELAN .....   | 29 |
| Hình 12: Tổng quan mô hình Vision Transformer .....  | 31 |
| Hình 13: Tổng quan kiến trúc mô hình RT-DETR .....   | 31 |
| Hình 14: Quá trình chuẩn bị tập dữ liệu .....  | 33 |
| Hình 15: Quá trình huấn luyện mô hình phát hiện đối tượng .....  | 33 |
| Hình 16: Quá trình huấn luyện mô hình phân lớp với các kiến trúc: LeNet, MobileNet, InceptionV3, VGG16, ResNet101, DenseNet201 ..... | 34 |
| Hình 17: Cấu trúc cây thư mục tập dữ liệu huấn luyện mô hình phát hiện nguyên liệu .....   | 34 |
| Hình 18: Cấu trúc cây thư mục tập dữ liệu huấn luyện mô hình phân lớp .....  | 35 |
| Hình 19: Xây dựng mô hình nhận diện các thành phần nguyên liệu .....   | 36 |
| Hình 20: Xây dựng mô hình phân lớp món ăn đặc sản .....  | 38 |
| Hình 21: Tiền xử lý và gán nhãn dữ liệu .....  | 42 |
| Hình 22:Cấu trúc mạng nơ-ron của mô hình Yolov8.....   | 43 |
| Hình 23: Kết quả đánh giá mô hình Yolov8 sau khi huấn luyện .....  | 44 |
| Hình 24: Cấu trúc mạng nơ-ron của mô hình RT-DETR .....  | 45 |
| Hình 25: Kết quả đánh giá mô hình RT-DETR sau khi huấn luyện .....   | 46 |
| Hình 26: Cấu trúc mạng của mô hình Yolov9.....   | 47 |
| Hình 27: Kết quả đánh giá mô hình Yolov9 sau khi huấn luyện .....  | 48 |

|  |    |
|--|----|
| Hình 28: So sánh các mô hình xác định món ăn đặc sản Việt Nam..... | 59 |
| Hình 29: Giao diện hệ thống .....                                  | 60 |
| Hình 30: Kết quả upload hình ảnh.....                              | 61 |
| Hình 31: Kết quả xác định món ăn .....                             | 61 |

**DANH MỤC BẢNG**

|   |    |
|---|----|
| Bảng 1: Bộ luật các thành phần nguyên liệu có trong các món ăn đặc sản Việt Nam ..... | 22 |
| Bảng 2: Bộ luật các thành phần nguyên liệu có trong các món ăn đặc sản Việt Nam ..... | 39 |
| Bảng 3: Thống kê số lượng mẫu trong tập dữ liệu .....                                 | 43 |
| Bảng 4: Cấu trúc mạng của LeNet .....   | 49 |
| Bảng 5: Cấu trúc mạng của MobileNet .....   | 49 |
| Bảng 6: Cấu trúc mạng của InceptionV3 .....   | 50 |
| Bảng 7: Cấu trúc mạng của VGG16.....  | 50 |
| Bảng 8: Cấu trúc mạng của ResNet101.....  | 51 |
| Bảng 9: Cấu trúc mạng của DenseNet201 .....   | 51 |
| Bảng 10: Độ chính xác mô hình phân loại món ăn đặc sản Việt Nam.....                  | 52 |
| Bảng 11: Các thư viện được sử dụng .....  | 56 |
| Bảng 12: Thống kê độ chính xác, dung lượng của ba mô hình.....                        | 57 |
| Bảng 13: Thống kê kiểm thử các mô hình phát hiện thành phần nguyên liệu .....         | 58 |
| Bảng 14: Độ chính xác mô hình phân loại món ăn đặc sản Việt Nam.....                  | 59 |

## **DANH MỤC CÁC TỪ VIẾT TẮT**

| <b>STT</b> | <b>Từ viết tắt</b> | <b>Ngôn ngữ</b> | <b>Diễn giải</b>  |
|------------|--------------------|-----------------|---|
| 1          | AI                 | Tiếng Anh       | Artificial Intelligence                                 |
| 2          | CNTT & TT          | Tiếng Việt      | Công nghệ thông tin và Truyền thông                     |
| 3          | EL                 | Tiếng Anh       | Elementary Logic  |
| 4          | BERT               | Tiếng Anh       | Bidirectional Encoder Representations from Transformers |
| 5          | phoBERT            | Tiếng Anh       | Pre-trained language models for Vietnamese              |
| 4          | YOLO               | Tiếng Anh       | You Only Look Once                                      |
| 5          | Macro              | Tiếng Anh       | Macronutrient   |
| 6          | CNN                | Tiếng Anh       | Convolutional Neural Network                            |
| 7          | RT-DETR            | Tiếng Anh       | Real Time Detection Transformer                         |
| 8          | set_detected       | Tiếng Anh       | Tập hợp các nguyên liệu mà mô hình nhận dạng được       |
| 9          | NCKH               | Tiếng Việt      | Nghiên cứu khoa học                                     |
| 10         | PANet              | Tiếng Anh       | Path Aggregation Network                                |
| 11         | PGI                | Tiếng Anh       | programmable gradient information                       |
| 12         | GELAN              | Tiếng Anh       | Generalized Efficient Layer Aggregation Network         |
| 13         | ViT                | Tiếng Anh       | Vision Transformer                                      |
| 14         | AIFI               | Tiếng Anh       | Attention based Intra Scale Feature Interaction         |
| 15         | CCFM               | Tiếng Anh       | CNN based Cross-scale Feature-fusion Module             |
| 16         | UI                 | Tiếng Anh       | User Interface  |
| 17         | CPU                | Tiếng Anh       | Central Processing Unit                                 |
| 18         | RAM                | Tiếng Anh       | Random Access Memory                                    |
| 19         | GPU                | Tiếng Anh       | Graphics Processing Unit                                |
| 20         | PIL                | Tiếng Anh       | Python Imaging Library                                  |

## TÓM TẮT

Ẩm thực Việt Nam là một phần không thể thiếu trong văn hóa và đời sống của người Việt Nam. Với sự phong phú, đa dạng về các thành phần nguyên liệu và hương vị đặc trưng, ẩm thực Việt Nam đã thu hút sự quan tâm, yêu thích và nguồn cảm hứng vô tận cho các đầu bếp và nhà nghiên cứu ẩm thực trên khắp thế giới.

Tuy nhiên, việc nhận dạng các món ăn đặc sản của Việt Nam có thể gặp khó khăn đối với những người ngoại quốc hoặc người không quen biết với ẩm thực địa phương. Đặc điểm và cách chế biến của từng món ăn thường mang những đặc trưng riêng biệt, gây khó khăn cho việc phân biệt các món ăn với nhau. Vì vậy, việc xây dựng một hệ thống trí tuệ nhân tạo (AI) có khả năng xác định món ăn đặc sản của Việt Nam sẽ giúp cho việc tìm hiểu và trải nghiệm ẩm thực địa phương trở nên dễ dàng hơn đối với mọi người, đặc biệt là đối với du khách quốc tế. Đồng thời, hệ thống này cũng có thể hỗ trợ trong việc bảo tồn và phát triển văn hóa ẩm thực của Việt Nam. Hệ thống AI này có thể trở thành một nguồn tài nguyên hữu ích cho những người yêu thích tìm hiểu, nghiên cứu về ẩm thực Việt Nam.

Để tài nghiên cứu của chúng tôi sẽ tập trung phát hiện và nhận dạng các đối tượng là các thành phần nguyên liệu trên ảnh chụp về món ăn. Từ các thành phần nguyên liệu được tìm thấy này, sẽ áp dụng vào một bộ luật và xác định được kết quả là tên món ăn đặc sản trong ảnh chụp. Nghiên cứu này sẽ vận dụng các kiến thức về thuật toán, học sâu (deep learning) và thị giác máy tính để giải quyết bài toán.

## ABSTRACT

Vietnamese cuisine is an integral part of the culture and life of the Vietnamese people. With its richness and diversity in ingredients and distinctive flavors, Vietnamese cuisine has attracted attention, love, and endless inspiration for chefs and culinary researchers worldwide.

However, identifying Vietnamese specialty dishes may pose challenges for foreigners or those unfamiliar with local cuisine. The characteristics and methods of preparation of each dish often have unique features, making it difficult to distinguish between them. Therefore, developing an artificial intelligence (AI) system capable of identifying Vietnamese specialty dishes would facilitate understanding and experiencing local cuisine for everyone, especially international tourists. Additionally, this system could also support the preservation and development of Vietnamese culinary culture. Such an AI system could become a valuable resource for those interested in exploring and researching Vietnamese cuisine.

Ours research topic will focus on detecting and identifying objects as ingredients in food photos. From these identified ingredients, a set of rules will be applied to determine the result, which is the name of the specialty dish in the photo. This research will apply knowledge of algorithms, deep learning, and computer vision to solve the problem.

## PHẦN GIỚI THIỆU

### 1. Đặt vấn đề

Ẩm thực Việt Nam là một phần không thể tách rời của văn hóa dân tộc, với những món ăn đặc sản như phở, bún bò Huế, bún cá, mì Quảng... Sự đa dạng và phong phú giữa các thành phần nguyên liệu, nguồn gốc xuất xứ lâu đời... các món ăn đặc sản này không chỉ mang trong mình giá trị văn hóa, đại diện cho ẩm thực địa phương, hương vị độc đáo mà còn là cầu nối giao thương văn hóa, kinh tế với các quốc gia trên thế giới. Ẩm thực Việt Nam đóng vai trò quan trọng trong việc lan tỏa hình ảnh văn hóa của đất nước ra thế giới, thu hút du khách và đóng góp vào sự phát triển kinh tế, du lịch của đất nước.

Nền ẩm thực nước Việt có văn hóa, địa lý rộng lớn, thường xuyên giao thoa với các quốc gia láng giềng, cộng với lịch sử phát triển lâu đời, thường xuyên cập nhật về các nguyên liệu, quá trình chế biến, tên gọi... Điều này có thể gây ra nhầm lẫn với du khách hoặc người Việt không quen thuộc với các món ăn Việt Nam. Trong quá trình du lịch và khám phá ẩm thực Việt Nam, không hiếm trường hợp du khách quốc tế hoặc người Việt không rành về tên gọi chính xác của các món ăn đặc sản của địa phương. Những hiểu lầm và nhầm lẫn thường xuyên xảy ra, khiến cho trải nghiệm ẩm thực trở nên phức tạp hơn. Những hiểu lầm này không chỉ gây ra sự bất tiện cho du khách mà còn khiến cho người bán hoặc người địa phương cảm thấy bất ngờ hoặc không hài lòng. Tuy nhiên, điều quan trọng nhất là sẵn lòng học hỏi và tôn trọng văn hóa ẩm thực của đất nước mình đang ghé thăm, bằng cách nỗ lực để hiểu rõ hơn về các món ăn và tên gọi chính xác của chúng.

Việc xây dựng một hệ thống trí tuệ nhân tạo (AI) có khả năng xác định các món ăn đặc sản Việt Nam là cực kỳ cần thiết trong bối cảnh ngành du lịch và ẩm thực đang phát triển mạnh mẽ. Một hệ thống như vậy không chỉ giúp du khách quốc tế hiểu rõ hơn về ẩm thực Việt Nam mà còn tạo điều kiện thuận lợi cho việc giới thiệu và quảng bá ẩm thực đặc sản của đất nước.

Bằng cách áp dụng trí tuệ nhân tạo, học sâu (deep learning), thị giác máy tính và thuật toán. Hệ thống AI có thể được huấn luyện và có khả năng nhận dạng các món ăn đặc sản dựa trên hình ảnh. Việc này sẽ giúp du khách dễ dàng tìm hiểu về tên gọi, nguyên liệu và cách chế biến của từng món ăn, giúp họ có trải nghiệm ẩm thực đầy đủ và đúng chất.

Hơn nữa, việc xây dựng một hệ thống AI xác định món ăn cũng sẽ hỗ trợ các doanh nghiệp trong ngành du lịch và nhà hàng. Họ có thể sử dụng công nghệ này để tạo ra các ứng dụng di động hoặc website cung cấp thông tin chi tiết về các món ăn đặc sản, đồng thời tạo ra trải nghiệm tương tác và hấp dẫn cho khách hàng.

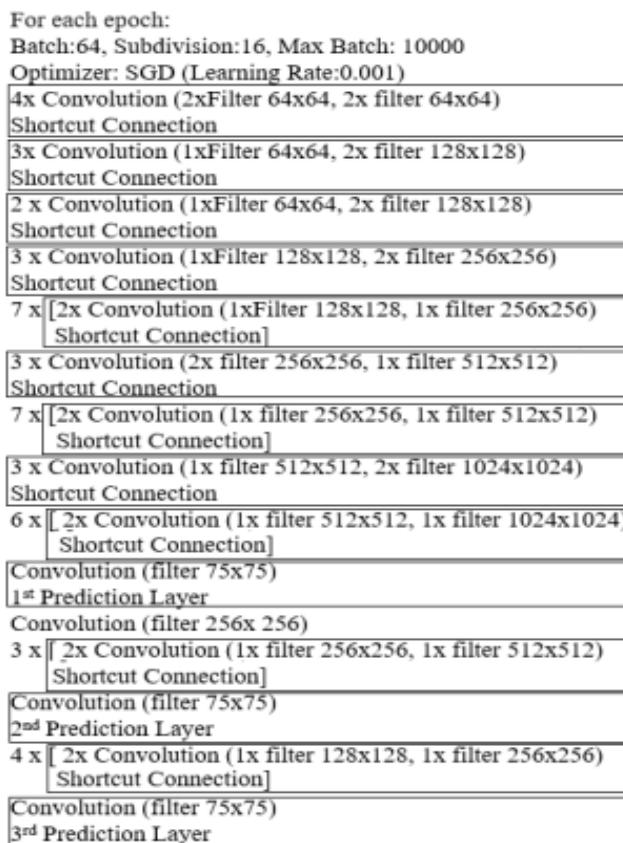
Ngoài ra, việc sử dụng công nghệ AI cũng sẽ giúp chính phủ và các tổ chức du lịch thu thập dữ liệu về ẩm thực Việt Nam một cách hiệu quả hơn, từ đó đề xuất các chính sách và chiến lược phát triển du lịch phù hợp. Đồng thời, việc quản lý và bảo tồn di sản ẩm thực cũng được thúc đẩy mạnh mẽ hơn.

Việc xây dựng một hệ thống AI xác định món ăn đặc sản Việt Nam không chỉ mang lại lợi ích cho du khách mà còn đóng góp vào sự phát triển bền vững của ngành du lịch và ẩm thực đất nước. Điều này là cực kỳ cần thiết trong bối cảnh hội nhập quốc tế ngày càng sâu rộng và nhu cầu khám phá văn hóa địa phương ngày càng tăng cao.

## 2. Lịch sử giải quyết vấn đề

### 2.1. A Real-time Junk Food Recognition System based on Machine Learning (2021)

Bài báo “A Real-time Junk Food Recognition System based on Machine Learning) (1) được lưu trữ tại ArXiv. Một nghiên cứu thuộc Department of Computer Science, American International University-Bangladesh, hình 1 được trích từ bài nghiên cứu, thể hiện kiến trúc nội bộ DarkNet53 của mô hình Yolov3 trong nghiên cứu của họ.



Hình 1: Kiến trúc DarkNet53 của mô hình Yolov3

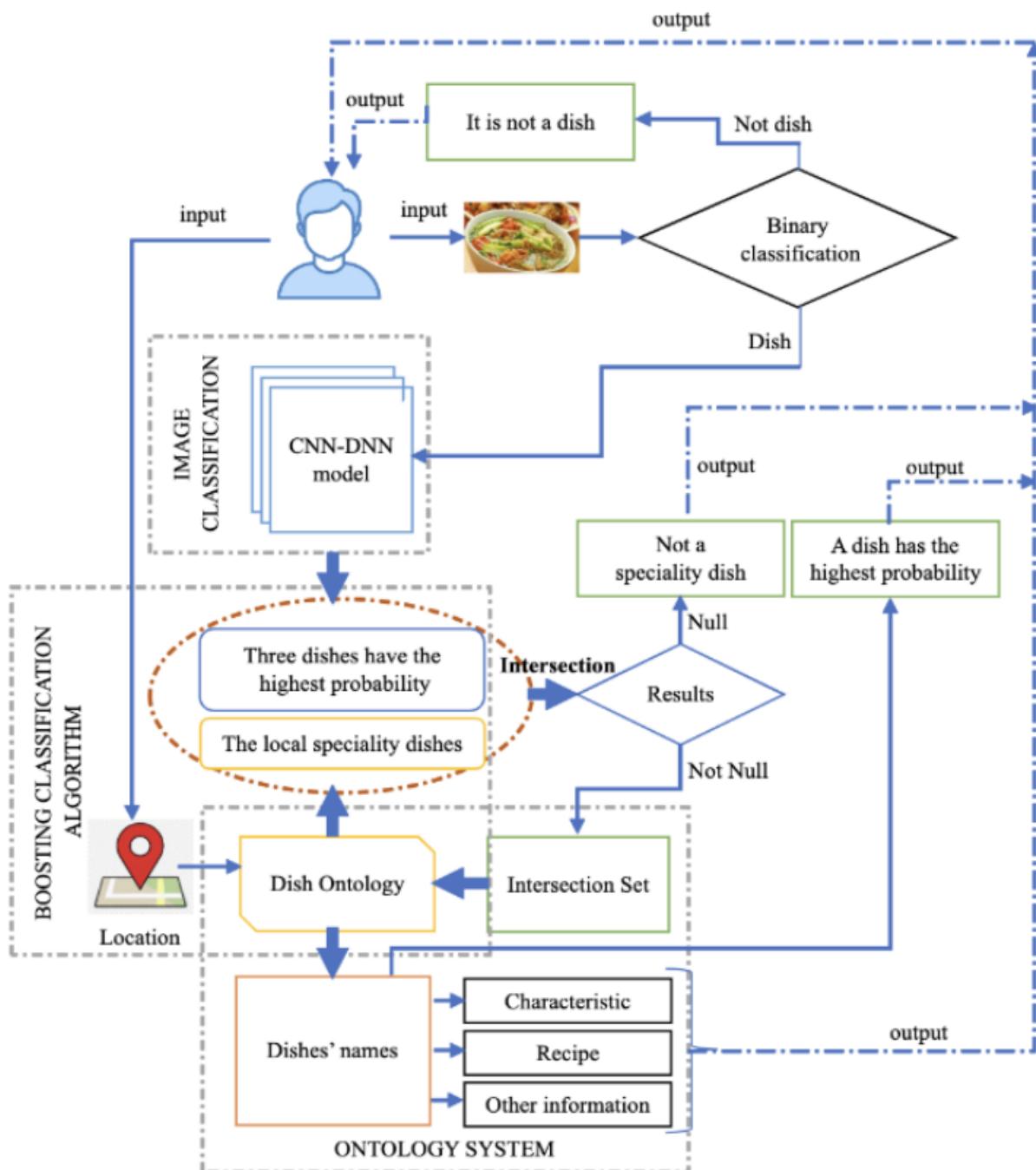
Mọi người luôn tìm kiếm thức ăn ngon miệng, với thực trạng tiêu thụ các loại thực phẩm không lành mạnh là nguồn phổ biến nhất. Điều này gây tổn hại tới sức khỏe của con người và là nguy cơ con người mắc các bệnh lý nguy hiểm. Phương pháp học máy (machine learning) được ứng dụng trong hầu hết cuộc sống của chúng ta, một trong số đó là nhận dạng đối tượng thông qua xử lý hình ảnh.

Trong đề tài nghiên cứu này, nhóm tác giả đã tạo ra một tập dữ liệu gồm 2.000 mẫu dữ liệu ban đầu từ 20 loại thức ăn không tốt cho sức khỏe như: burger, hotdog, pizza, bánh phô mai... để thử tạo ra mô hình có khả năng nhận dạng thực phẩm không lành mạnh. Sau quá trình huấn luyện, mô hình cho ra kết quả có độ chính xác là 61.6%. Quá trình tăng cường dữ liệu (data augmentation) được tiến hành, áp dụng các phương pháp gồm: xoay ảnh với các góc độ ngẫu nhiên, lật ngang ảnh, lật dọc ảnh, thu phóng lại tỷ lệ. Sau quá trình tăng cường dữ liệu, mỗi lớp thu được 500 ảnh, tổng cộng tập dữ liệu có 10.000 mẫu.

Nhóm tác giả đã sử dụng framework YOLO, cụ thể là mô hình Yolov3, kiến trúc Darknet-53 để huấn luyện mô hình. Kết quả huấn luyện cho ra mô hình với độ chính xác là 98.05%.

## **2.2. Xác định món ăn đặc sản Việt Nam với sự kết hợp của mạng học sâu và bản thể học (2023)**

Bài báo chính thức được đăng ở Tạp chí Khoa học Đại học Cần Thơ. Đây là một nghiên cứu thuộc khoa Khoa học máy tính, trường CNTT & TT, trường Đại học Cần Thơ (2). Hình 2 được trích trong bài báo, trình bày mô hình cho bảo tồn món ăn đặc trưng Việt Nam được nhóm tác giả đề xuất.



Hình 2: Mô hình cho bảo tồn món ăn đặc trưng Việt Nam

Nhận thấy hiện nay có rất ít nghiên cứu và ứng dụng trí tuệ nhân tạo (AI) tập trung vào lĩnh vực bảo tồn và phổ biến các giá trị của truyền thống văn hóa ẩm thực, hầu hết những nghiên cứu chỉ tập trung vào phân lớp hình ảnh và thiếu thông tin toàn diện của từng món ăn.

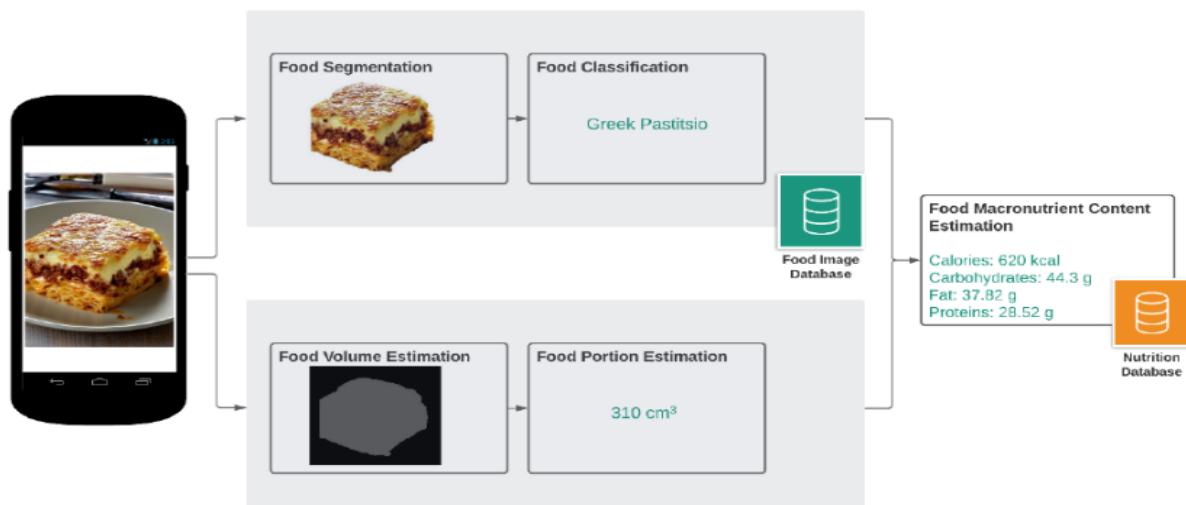
Đội ngũ tham gia nghiên cứu đã đề xuất về khung AI cho các món ăn Việt Nam được giới thiệu trong bài báo này. Cụ thể, một bản thẻ luận (ontology) món ăn đặc sản Việt Nam để lưu trữ thông tin liên quan và mô hình phân lớp hình ảnh các món ăn đặc sản được trình bày. Đóng góp chính của bài báo bao gồm:

1. Mô hình phân lớp đặc sản ẩm thực Việt Nam ở đồng bằng sông Cửu Long phục vụ du lịch, với độ chính xác trên 96%;
2. Bản thể luận đặc sản ẩm thực Việt Nam;
3. Bộ dữ liệu hình ảnh ẩm thực Việt Nam được thu thập;
4. Ứng dụng AI để khám phá ẩm thực ở đồng bằng sông Cửu Long.

Một bản thể luận nhẹ (về các món ăn truyền thống của Việt Nam) được phân lớp theo vùng và tỉnh. Việc triển khai bản thể luận được mã hóa bằng logic mô tả EL. Tận dụng một mạng tích chập sâu để phát hiện và nhận ra tên món ăn. Để tăng độ chính xác của phân lớp, lý do lựa chọn ba món ăn có độ chính xác tốt nhất để tiến hành lựa chọn duy nhất thông qua bản thể học. Cụ thể, tận dụng khả năng lập luận (lý luận) của bản thể học để đưa ra quyết định với một số thông tin bổ sung hữu ích, tức là vị trí của món ăn và mô tả các đặc điểm của nó. Cuối cùng, một truy vấn thông tin bằng SPARQL để có tên thực phẩm cuối cùng được thực hiện. Sau khi có tên món ăn, hệ thống cung cấp thêm thông tin chi tiết về nhà cung cấp, cách chế biến, video nấu ăn và nguyên liệu. Ngoài ra, hệ thống còn cung cấp ứng dụng chatbot để hỗ trợ khách du lịch tìm kiếm thông tin. Để xây dựng ứng dụng này, mô hình BERT và mô hình PhoBERT được sử dụng. Một điểm cần lưu ý là nghiên cứu này tập trung chủ yếu vào việc đề xuất các ứng dụng thực tế (một khung) có thể áp dụng tốt cho du lịch và bảo tồn văn hóa truyền thống ở Việt Nam.

### **2.3. A Review of Image-based Food Recognition and Volume Estimation Artificial Intelligence Systems (2023)**

Bài báo được lưu trữ tại ResearchGate (3). Nghiên cứu này tận dụng hai bài toán phổ biến trong thị giác máy tính gồm: phân đoạn đối tượng (object segmentation) và phân lớp đối tượng (object classification). Hình 3 được trích từ bài báo, mô tả quy trình hoạt động của hệ thống đánh giá chế độ dinh dưỡng.



*Hình 3: Mô tả quy trình hoạt động của hệ thống đánh giá chế độ dinh dưỡng*

Chế độ ăn uống lành mạnh hàng ngày và hấp thụ cân bằng các chất dinh dưỡng thiết yếu đóng một vai trò quan trọng trong lối sống hiện đại. Việc ước tính hàm lượng chất dinh dưỡng có trong bữa ăn là một việc không thể thiếu để duy trì sức khỏe, đặc biệt là ở các bệnh nhân mắc các chứng bệnh nghiêm trọng, chẳng hạn như tiểu đường, béo phì và bệnh tim mạch. Việc ước tính lượng thành phần các chất dinh dưỡng có trong bữa ăn một cách tự động, chính xác và thời gian thực, có thể được giải quyết bằng thị giác máy tính, bằng cách sử dụng hình ảnh món ăn được chụp thông qua điện thoại của người dùng.

Trong bài nghiên cứu trên, nhóm tác giả sẽ tiến hành phân đoạn hình ảnh món ăn (food image segmentation), đồng thời ước tính thể tích của món ăn. Đôi tượng sau khi phân đoạn sẽ được chuyển đến giai đoạn phân lớp, kết quả phân lớp sẽ được truy vấn trong cơ sở dữ liệu hình ảnh món ăn của hệ thống. Từ thể tích và kết quả phân lớp của món ăn sẽ tính được lượng Macro – là ba thành phần dinh dưỡng có trong bữa ăn bao gồm: protein (chất đạm), carbohydrate (tinh bột), fat (chất béo). Từ ba thành phần dinh dưỡng này sẽ tính được lượng calories (năng lượng calo) của bữa ăn.

### 3. Mục tiêu đề tài

#### 3.1. Mục tiêu chung

Xây dựng hệ thống nhận dạng món ăn đặc sản Việt Nam dựa trên các thành phần nguyên liệu, cụ thể là vận dụng các kiến thức về lĩnh vực học máy, học sâu, thị giác máy tính. Thông qua kết quả nhận dạng là tập hợp thành phần nguyên liệu, có thể thực hiện xác định được tên gọi của món ăn. Cuối cùng, một website sẽ được cài đặt để người dùng cuối có thể sử dụng mô hình AI này. Đề tài được chia ra làm ba giai đoạn gồm:

1. Huấn luyện mô hình phát hiện đối tượng (object detection);
2. Thiết kế thuật toán dùng để xác định tên gọi món ăn;
3. Triển khai mô hình thành ứng dụng cho người dùng cuối sử dụng.

Lưu ý, trong đề tài này chúng tôi sẽ tập trung vào giai đoạn hai để thiết kế một thuật toán mới có khả năng xác định được đối tượng dựa trên một tập hợp các chi tiết về đối tượng đó. Đây sẽ là khởi đầu cho các nghiên cứu tiếp theo để phát triển và nâng cao độ chính xác của các giải thuật học máy.

### **3.2. Mục tiêu chi tiết**

1. Thu thập và tiền xử lý dữ liệu hình ảnh để xây dựng một tập dữ liệu hình ảnh về các món ăn đặc sản của Việt Nam;
2. Gán nhãn dữ liệu cho tập dữ liệu hình ảnh ở mục tiêu 1, để chuẩn bị dữ liệu cho giai đoạn huấn luyện mô hình nhận dạng các đối tượng nguyên liệu có trong ảnh món ăn;
3. Soạn ra một bộ luật bao gồm tên món ăn, các thành phần nguyên liệu tương ứng của món ăn đó;
4. Xây dựng công thức dùng để tính toán tên gọi của món ăn dựa trên tập nguyên liệu được AI nhận dạng và bộ luật đã soạn;
5. Huấn luyện các mô hình phát hiện đối tượng gồm: RT-DETR, Yolov8, Yolov9;
6. Tìm hiểu các phương pháp huấn luyện, các kỹ thuật tối ưu trong quá trình huấn luyện nhằm tăng tính chính xác cho mô hình;
7. Đánh giá độ chính xác của các mô hình nhận dạng nguyên liệu;
8. Tạo thêm tập dữ liệu phục vụ quá trình huấn luyện các mô hình phân lớp (classification) món ăn CNN;
9. Huấn luyện các mô hình phân lớp món ăn CNN với các kiến trúc khác nhau bao gồm: LeNet, MobileNet, InceptionV3, VGG16, ResNet101, DenseNet201;
10. So sánh độ chính xác của các mô hình phân lớp món ăn với mô hình nhận dạng thành phần nguyên liệu;
11. Triển khai mô hình lên website để người dùng cuối có thể sử dụng.

## **4. Phạm vi nghiên cứu**

### **4.1. Đối tượng nghiên cứu**

- Các món ăn đặc sản của nước Việt Nam gồm: bún cá, hủ tiếu Mỹ Tho, bún nước lèo, cơm tấm Long Xuyên, bún hải sản bè bè, bánh hỏi heo quay, cơm gà, cao lầu, mì Quảng, bún bò Huế, phở Hà Nội, bún mực, bún mọc, bún đậu mắm tôm;
- Xác định tên gọi món ăn dựa vào tập hợp nguyên liệu;
- Các công thức đo khoảng cách từ phần tử mới đến các đối tượng hiện có;
- Các mô hình phát hiện đối tượng gồm RT-DETR, Yolov8, Yolov9;
- Các kiến trúc học sâu (deep learning) gồm: LeNet, MobileNet, InceptionV3, VGG16, ResNet101, DenseNet201.

#### **4.2. Phạm vi nghiên cứu**

- Các món ăn đặc sản của nước Việt Nam;
- Thực hiện với các món ăn có nhiều chi tiết nguyên liệu;
- Thực hiện trên ảnh chụp 2D;
- Xác định món ăn bằng phương pháp tính khoảng cách;
- Xây dựng các mô hình được áp dụng các phương pháp học sâu.

#### **5. Phương pháp nghiên cứu**

##### **5.1. Phương pháp nghiên cứu tài liệu**

Khảo sát các phương pháp giải quyết bài toán có liên quan đến đề tài nghiên cứu về phát hiện và đo lường khoảng cách đối tượng. Cụ thể, chúng tôi tập trung chủ yếu với bao gồm hai phần chính: kiến thức được tổng hợp lại từ các bài báo, nghiên cứu phân tích ẩm thực Việt Nam và kiến thức trí tuệ nhân tạo để xây dựng mô hình.

- ❖ **Kiến thức tổng hợp được từ bài báo, nghiên cứu về ẩm thực Việt Nam:** tham khảo các thành phần nguyên liệu được cho rằng là phổ biến đối với món ăn đặc sản đó sẽ gồm những nguyên liệu nào.
- ❖ **Kiến thức Trí tuệ nhân tạo:** các mô hình phát hiện đối tượng để thu được tập hợp các nguyên liệu có trong ảnh chụp món ăn. Từ đó, nhận xét, đánh giá, so sánh các mô hình và chọn ra mô hình có độ chính xác tốt nhất. Tìm hiểu phương pháp huấn luyện mô hình, tối ưu trong quá trình huấn luyện và các phương pháp đánh giá mô hình.

##### **5.2. Phương pháp nghiên cứu thực nghiệm**

Xây dựng bộ dữ liệu phục vụ cho quá trình huấn luyện mô hình. Thu thập dữ liệu hình ảnh bằng cách chụp ảnh món ăn ngoài thực tế kết hợp với thu thập hình ảnh món ăn trên Google Images. Sau khi thu thập dữ liệu sẽ tiến hành gán nhãn dữ liệu. Gán nhãn hoàn thành, phân chia tập dữ liệu ra ba phần gồm: tập huấn luyện (training), tập xác thực (validation), tập kiểm tra (testing). Sau đó sẽ đưa tập dữ liệu vào huấn luyện mô hình nhận dạng nguyên liệu. Cài đặt môi trường huấn luyện, tinh chỉnh các tham số, kỹ thuật để nâng cao kết quả của mô hình đề xuất. Nhận xét đánh giá mô hình dựa vào tiêu chí độ chính xác. Chạy thực nghiệm mô hình trên ứng dụng web.

## 6. Kết quả đạt được

Đề tài đã vận dụng các thuật toán trí tuệ nhân tạo vào thực tế. Cụ thể, xây dựng một hệ thống có khả năng nhận dạng món ăn đặc sản Việt Nam dựa trên các chi tiết nguyên liệu. Những đóng góp chính của nghiên cứu như sau:

1. Tập dữ liệu gồm các món ăn đặc sản Việt Nam;
2. Bộ luật các thành phần nguyên liệu của các món ăn đặc sản Việt Nam;
3. Xây dựng ba mô hình có khả năng nhận dạng các thành phần nguyên liệu có trên ảnh chụp món ăn gồm: RT-DETR, Yolov8, Yolov9;
4. Đóng góp một công thức tính toán mới dùng để xác định được một đối tượng dựa trên các chi tiết về đối tượng đó;
5. Xây dựng được một website, nơi người dùng có thể sử dụng mô hình để xác định món ăn đặc sản Việt Nam.

## 7. Bố cục bài báo cáo

Nội dung của quyển báo cáo niêm luân gồm các phần sau đây:

### Phần giới thiệu

Phần này trình bày các vấn đề phát sinh và lịch sử giải quyết vấn đề của đề tài, mục tiêu đề tài, những nghiên cứu được thực hiện trong lúc thực hiện đề tài.

### Phần nội dung

Phần này trình bày chi tiết bài toán, thiết kế và cài đặt hệ thống, đồng thời nêu lên quy trình kiểm thử, đánh giá mô hình. Bao gồm các phần:

**Chương 1:** Mô tả bài toán.

**Chương 2:** Thiết kế, cài đặt giải thuật, trình bày các bước xây dựng hệ thống.

**Chương 3:** Kiểm thử hệ thống và đánh giá độ chính xác.

### Phần kết luận

Phần này trình bày kết quả đạt được của đề tài cũng như những hạn chế mà đề tài chưa thực hiện được, ngoài ra cũng đưa ra hướng phát triển sau này.

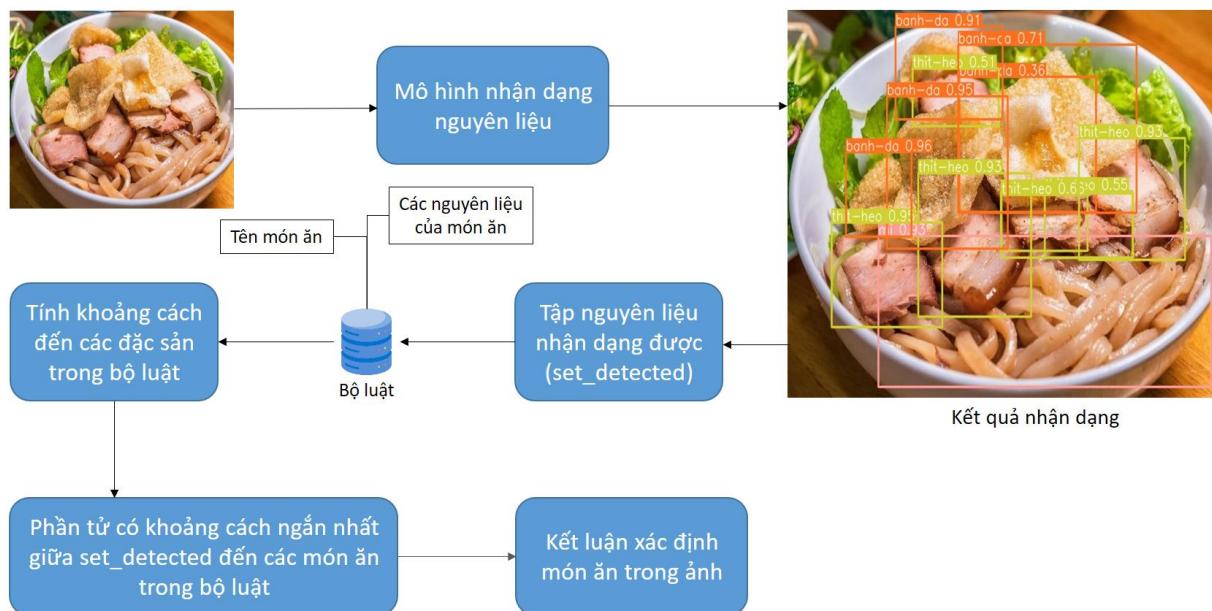
## PHẦN NỘI DUNG

### CHƯƠNG 1 MÔ TẢ BÀI TOÁN

#### 1.1. Mô tả chi tiết bài toán

Đề tài “**Xây dựng hệ thống xác định món ăn đặc sản Việt Nam dựa trên các thành phần nguyên liệu**” được nghiên cứu nhằm đáp ứng nhu cầu tìm hiểu về món ăn trong trường hợp người dùng chưa biết hoặc không rành về món ăn đó. Đối tượng người sử dụng hệ thống là toàn thể tất cả mọi người, đặc biệt tập trung vào đối tượng là các du khách trong và ngoài nước, những người không rành về ẩm thực Việt và các món ăn địa phương. Hệ thống sẽ có tính ứng dụng cao trong công tác du lịch, giúp trải nghiệm du lịch của du khách trở nên thú vị hơn, đóng góp quảng bá hình ảnh ẩm thực Việt Nam ra quốc tế một cách mạnh mẽ.

Quy trình hoạt động của hệ thống, đầu tiên, người dùng sẽ chụp ảnh về món ăn đặc sản, hình ảnh sẽ được tải lên hệ thống. Mô hình AI sẽ tiến hành đọc ảnh đầu vào, và phát hiện, nhận dạng các đối tượng nguyên liệu trên hình ảnh. Tập hợp nguyên liệu nhận dạng được (`set_detected`) sẽ được sử dụng làm đầu vào cho thuật toán tính khoảng cách giữa `set_detected` với các món ăn có trong dữ liệu (bộ luật) món ăn đặc sản của hệ thống. Sau khi tính được khoảng cách giữa `set_detected` đến các món ăn đặc sản trong bộ luật, đặc sản có khoảng cách ngắn nhất sẽ là kết quả xác định món ăn của hệ thống. Hình 4 thể hiện quy trình hoạt động của hệ thống.



Hình 4: Quy trình xác định món ăn của hệ thống

## 1.2. Vấn đề và giải pháp liên quan đến bài toán

### 1.2.1. Các nguyên liệu có trong món ăn đặc sản Việt Nam

Ẩm thực Việt Nam không chỉ là nơi tìm thấy các món ăn ngon và phong phú mà còn là nền văn hóa ẩm thực phản ánh sự kết hợp tinh tế giữa các nguyên liệu để tạo ra hương vị đặc trưng. Ví dụ, trong món phở, sự kết hợp giữa nước dùng từ xương gà hoặc bò, thịt, các loại gia vị như gừng, hành, hạt tiêu và rau sống như húng quế, ngò gai, rau mùi. Bún cá thường được làm từ bún, cùng với nước dùng được nấu từ xương cá hoặc hải sản khác như tôm, cua, hay cá lóc. Cá được cắt thành từng miếng nhỏ và nấu chín mềm trong nước dùng, được phục vụ cùng với tôm và các loại rau như rau răm, lá chuối, lá lốt.

Mỗi vùng miền của Việt Nam có những nguyên liệu đặc trưng riêng, phản ánh điều kiện tự nhiên và văn hóa của địa phương đó. Ví dụ, miền Trung có nhiều món sử dụng các loại tôm, cá, sò điệp từ biển, trong khi miền Nam thường sử dụng nhiều loại rau cải và đậu, còn miền Bắc có nhiều món ăn sử dụng thịt lợn, gà, và các loại rau sống. Mỗi vùng miền, mỗi địa phương sẽ có xu hướng ăn uống riêng kết hợp với đa dạng nguồn nguyên liệu, dẫn đến nét riêng trong công đoạn sử dụng các thành phần nguyên liệu, góp phần tạo nên sự đa dạng cho cách thức chế biến món ăn. Hình 5 thể hiện tuy cùng một món bún cá nhưng khi mang về mỗi tỉnh thành lại biến hóa thành một nét đặc trưng riêng của tỉnh thành đó.



Bún cá An Giang



Bún cá Kiên Giang



Bún cá Nha Trang



Bún cá Quy Nhơn



Bún cá Hà Nội



Bún cá Thái Bình

Hình 5: Món bún cá ở một số tỉnh thành của Việt Nam

Các món ngon đặc sản của Việt Nam là chủ đề được nghiên cứu, phân tích, nhận xét qua các bài báo, nghiên cứu trong nước và quốc tế. Ví dụ một bài báo nghiên cứu khoa học “Một số món ngon đặc sản của các tỉnh Đồng bằng Sông Cửu Long”, Tạp chí NCKH và Phát triển kinh tế Trường Đại học Tây Đô (4), hình 6 mô tả một số món ngon đặc sản mà nhóm tác giả có thực hiện phân tích. Chúng tôi đã tổng hợp thông tin từ nhiều bài báo và nghiên cứu tương tự, từ đó hiểu hơn về văn hóa ẩm thực nước nhà và xây dựng được một bộ luật các thành phần nguyên liệu có trong món ăn đặc sản như bảng 1.



Bánh pía



Bánh xèo



Bún măng vịt



Bánh canh ghe



Hủ tiếu Nam Vang



Chè trôi nước

*Hình 6: Một số món ăn đặc sản Việt Nam*

*Bảng 1: Bộ luật các thành phần nguyên liệu có trong các món ăn đặc sản Việt Nam*

| Món ăn             | Nguyên liệu ưu tiên | Nguyên liệu chính                           | Nguyên liệu phụ            | Tỉnh thành         |
|--------------------|---------------------|---|----------------------------|--------------------|
| Bún cá             | bún                 | cá  | tôm, rau răm               | An Giang           |
| Hủ tiếu Mỹ Tho     | hủ tiếu             | thịt băm, thịt heo, gan                     | trứng, tôm                 | Tiền Giang         |
| Bún nước lèo       | bún                 | cá, tôm, thịt heo quay                      | rau răm                    | Sóc Trăng          |
| Cơm tấm Long Xuyên | cơm                 | sườn, bì, trứng                             | dưa chua                   | An Giang           |
| Bún hải sản bè bè  | bún                 | tôm tí                                      | tôm, mực                   | Vũng Tàu           |
| Bánh hỏi heo quay  | bánh hỏi            | thịt heo quay                               |                            | Cần Thơ            |
| Cơm gà             | cơm                 | thịt gà                                     | rau răm                    | Quảng Nam          |
| Cao lầu            | mì, thịt heo        | bánh đa                                     |                            | Quảng Nam          |
| Mì Quảng           | mì                  | bánh đa, trứng                              | thịt gà, thịt heo, cá, tôm | Quảng Nam, Đà Nẵng |
| Bún bò Huế         | bún                 | thịt bò                                     |                            | Thừa Thiên Huế     |
| Phở Hà Nội         | phở                 | thịt bò                                     |                            | Hà Nội             |
| Bún mực            | bún                 | mực   | tôm                        | Phú Yên            |
| Bún mọc            | bún                 | viên mọc                                    | thịt heo                   | Hà Nội             |
| Bún đậu mắm tôm    | bún                 | chả cόm, dồi sụn, đậu hũ, thịt heo, chả giò |                            | Hà Nội             |

Gọi tập hợp chứa các thành phần nguyên liệu mà mô hình AI đã nhận dạng được là “**món ăn X**”. Bộ luật ở bảng 1 sẽ được sử dụng để tính khoảng cách từ món ăn X đến các món ăn đặc sản được trình bày trong bộ luật (bún cá, hủ tiếu Mỹ Tho, bún nước lèo, cơm tấm Long Xuyên, bún hải sản bè bè...). Phần tử món đặc sản có giá trị khoảng cách gần nhất với món ăn X, phần tử này sẽ là đáp án tên món ăn của hệ thống.

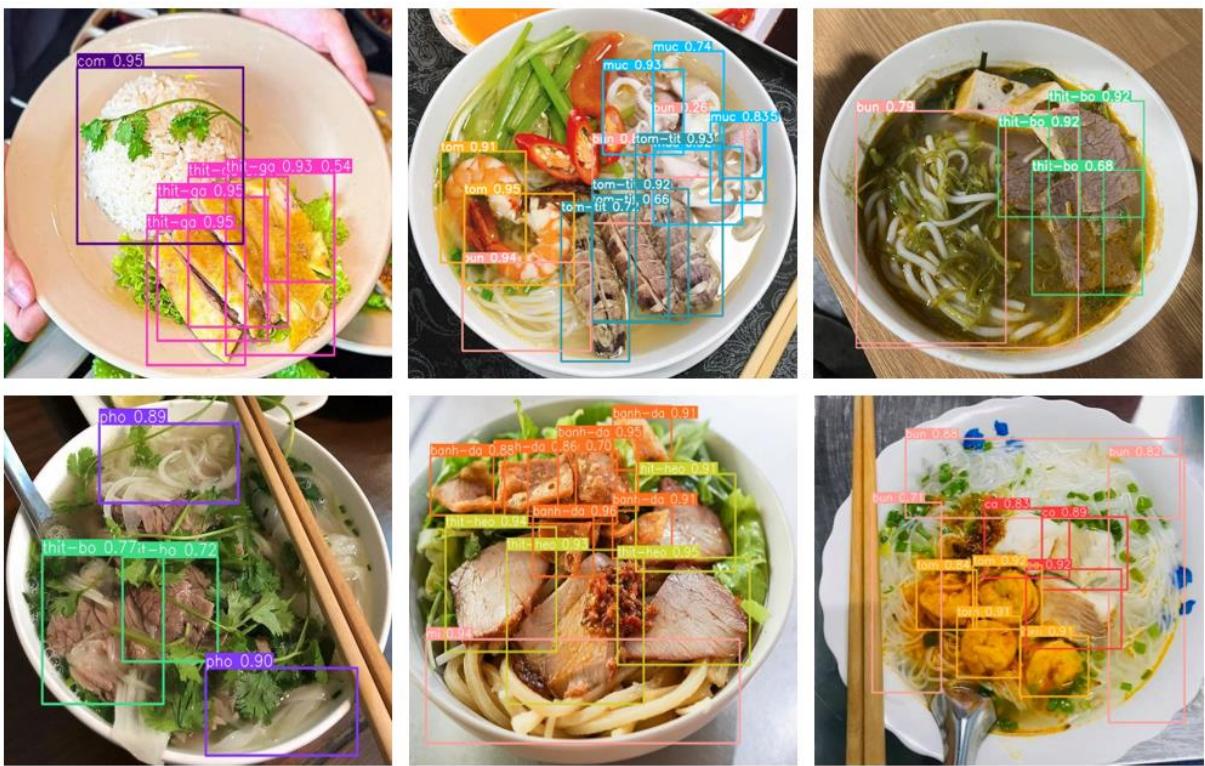
\*Lưu ý: Công thức tính khoảng cách giữa món ăn X và các món ăn đặc sản trong bộ luật sẽ được trình bày ở chương 2. Thiết và cài đặt giải thuật.

Cột “Món ăn” là tên gọi của các món ăn đặc sản, lần lượt các cột “Nguyên liệu ưu tiên”, “Nguyên liệu chính”, “Nguyên liệu phụ” là ba cột chứa thành phần nguyên liệu có trong món ăn. Các nguyên liệu được chia làm ba cấp độ, thể hiện tính chất quan trọng của nguyên liệu đó trong món ăn, từ cao đến thấp lần lượt là cột “Nguyên liệu ưu tiên”, “Nguyên liệu chính”, “Nguyên liệu phụ”. Cột “Tỉnh thành” thể hiện tên tỉnh thành được cho là nguồn gốc của các món ăn đặc sản. Nguồn thông tin của các tỉnh thành này được tổng hợp từ các tài liệu đáng tin cậy, nghiên cứu về ẩm thực Việt Nam.

### **1.2.2. Object detection**

Phát hiện đối tượng (object detection) là một kỹ thuật trong thị giác máy tính sử dụng mạng nơ-ron, giúp nhận diện và xác định vị trí của các đối tượng có trong hình ảnh. Hệ thống phát hiện đối tượng sẽ trả về tọa độ của các đối tượng trong ảnh mà nó đã được huấn luyện để nhận dạng. Hệ thống cũng sẽ trả về mức độ tin cậy, chỉ số này thể hiện mức độ tin cậy của hệ thống đối với dự đoán là chính xác. Mục tiêu của việc phát hiện đối tượng là phát triển các mô hình và kỹ thuật tính toán cung cấp một trong những kiến thức cơ bản nhất cần có cho các ứng dụng thị giác máy tính: *Đối tượng là gì và ở đâu?*

Phát hiện đối tượng phục vụ nhiều cho các công việc liên quan đến thị giác máy tính như phân đoạn đối tượng (object segmentation), chú thích hình ảnh (image captioning), theo dõi đối tượng (object tracking), ... Trong những năm gần đây, từ sự phát triển nhanh chóng của các kiến trúc học sâu, đã thúc đẩy sự tiến bộ của các mô hình phát hiện đối tượng. Phát hiện đối tượng trở thành đề tài được quan tâm, nghiên cứu nhiều hơn. Giờ đây, phát hiện đối tượng đã được sử dụng trong nhiều ứng dụng thực tế như hệ thống lái xe tự động, giám sát video... Hình 7 thể hiện một số kết quả mà chúng tôi đã huấn luyện mô hình Yolov8 để nhận diện thành phần nguyên liệu trên món ăn.



Hình 7: Một số kết quả nhận diện nguyên liệu của mô hình Yolov8

Xem xét hình 7 các đối tượng nguyên liệu món ăn được AI nhận dạng và xác định vị trí bằng các khung hình chữ nhật bao lấy đối tượng, khung hình chữ nhật này gọi là **bounding box**. Mỗi bounding box sẽ đi kèm với tên mà mô hình đã phân lớp của đối tượng, cùng với độ tin cậy (confidence) của đối tượng đó.

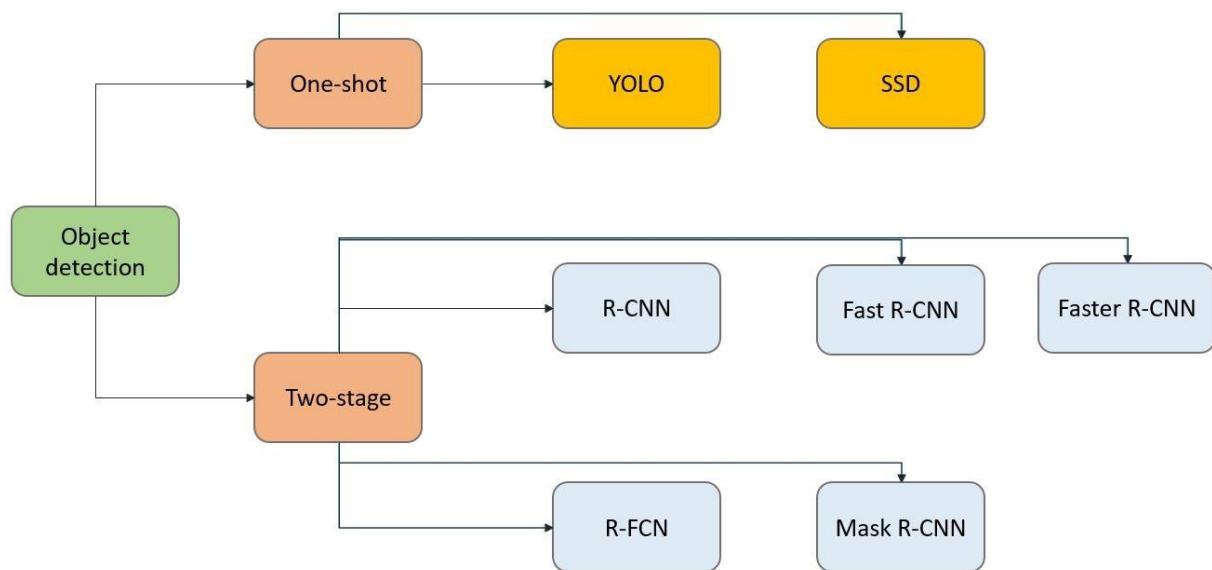
Phương pháp phát hiện đối tượng có thể được chia thành hai phương pháp tiếp cận bao gồm: dựa trên học máy và học sâu. Trong các phương pháp truyền thống dựa trên học máy, các kỹ thuật thị giác máy tính được sử dụng để xem xét các đặc trưng khác nhau của hình ảnh, như biểu đồ màu sắc hoặc các cạnh, để xác định các nhóm pixel có thể thuộc về một đối tượng. Các đặc điểm này sau đó được đưa vào một mô hình hồi quy để dự đoán vị trí của đối tượng trong hình ảnh cùng với nhãn của đối tượng đó. Trong khi đó, cách tiếp cận dựa trên học sâu sử dụng mạng nơ-ron tích chập (CNN) (5) để thực hiện phát hiện đối tượng không giám sát bằng cách sử dụng một quy trình end-to-end. Trong quy trình này, mạng CNN được huấn luyện để tự động học các đặc điểm quan trọng từ dữ liệu hình ảnh mà không cần phải xác định và trích xuất các đặc trưng riêng biệt trước.

CNN là một loại kiến trúc mạng nơ-ron được thiết kế để nhận diện các mẫu trong dữ liệu không gian, như hình ảnh. Chúng sử dụng các lớp tích chập để tự động học các bộ lọc có thể nhận diện các đặc điểm cụ thể trong ảnh, như cạnh, góc, hoặc các đặc trưng phức tạp hơn của các đối tượng. Quá trình huấn luyện của CNN thường bao gồm việc cung cấp cho mạng một lượng lớn các hình ảnh đã được gán nhãn (đối với phát hiện đối tượng, các hình ảnh này sẽ có các vị trí và nhãn của các đối tượng trong đó). CNN sau đó tự động điều chỉnh các tham số của nó để tối ưu hóa khả năng dự đoán vị trí và nhãn của các đối tượng trên các hình ảnh mới mà nó chưa từng thấy trước đó. Điểm mạnh của phương pháp này là tính tự động và tự học của nó. Thay vì phải xác định và chọn lọc các đặc điểm quan trọng thủ công, CNN tự động học các đặc trưng từ dữ liệu, giúp cho việc phát hiện đối tượng trở nên hiệu quả hơn và linh hoạt hơn trong nhiều tình huống khác nhau. Các mô hình đại diện cho phương pháp sử dụng CNN để phân lớp đối tượng và mạng hồi quy để dự đoán tọa độ bounding box như R-CNN, Fast R-CNN, Faster R-CNN và YOLO.

### 1.2.3. YOLO

YOLO (6) là viết tắt của “You Only Look Once”, được phát triển vào năm 2016 bởi Joseph Redmon và các đồng nghiệp. YOLO là một mô hình phát hiện đối tượng phổ biến được biết đến nhờ vào ưu điểm về tốc độ (thực hiện trong thời gian thực – real time) và độ chính xác.

Khi tiếp cận phương pháp phát hiện đối tượng dựa trên học sâu, thuật toán phát hiện đối tượng có thể được chia thành hai loại: bộ phát hiện một lần (single-shot detectors) và bộ phát hiện hai giai đoạn (two-stage detectors). Hình 8 là sơ đồ thể hiện hai loại thuật toán phát hiện đối tượng và một số mô hình đại diện tương ứng với từng loại.



Hình 8: Phân loại mô hình object detection

Phương pháp phát hiện đối tượng một lần (Single-shot object detection) sử dụng một lần duyệt đi qua hình ảnh đầu vào để dự đoán về sự hiện diện và vị trí của các đối tượng trong hình ảnh. Nó xử lý toàn bộ hình ảnh trong một lần duyệt, làm cho chúng hiệu quả về mặt tính toán. YOLO là một single-shot detector sử dụng mạng CNN để xử lý ảnh.

Thuật toán YOLO sẽ chia hình ảnh thành một lưới, một ô trong lưới dự đoán một hộp giới hạn (bounding box) nhất định. Ứng với mỗi bounding box, ô cũng dự đoán xác suất của lớp, cho biết khả năng một đối tượng cụ thể có xuất hiện trong bounding box. Quá trình nhận dạng bounding box trong YOLO bao gồm các bước sau:

- 1. Tạo lưới:** Hình ảnh được chia thành lưới SxS. Nếu tâm của một đối tượng rơi vào một ô lưới thì ô lưới đó có nhiệm vụ phát hiện đối tượng đó;
- 2. Dự đoán bounding box:** Mỗi ô lưới dự đoán B bounding box và độ tin cậy (confidence) cho các hộp đó. Độ tin cậy phản ánh mức độ chắc chắn của mô hình rằng bounding box này chứa một đối tượng. Đồng thời, độ tin cậy này cũng thể hiện độ chính xác của bounding box được mô hình dự đoán;
- 3. Dự đoán xác suất lớp:** mô hình dự đoán xác suất cho mỗi lớp của đối tượng trong bounding box.

Ở bài nghiên cứu này, chúng tôi sẽ sử dụng mô hình Yolov8 và Yolov9, là hai mô hình mới nhất thuộc họ mô hình YOLO để phục vụ cho việc nhận dạng các thành phần nguyên liệu có trong món ăn.

#### 1.2.4. Yolov8

Yolov8<sup>1</sup> là một mô hình phát hiện đối tượng được phát triển bởi nhóm Ultralytics<sup>2</sup>, mô hình này nổi tiếng nhờ vào ưu điểm ở tốc độ và độ chính xác. Yolov8 được xây dựng trên dựa trên sự thành công của những phiên bản trước, giải quyết những hạn chế của các mô hình phiên bản trước và kết hợp các kỹ thuật tiên tiến để nâng cao hiệu suất. Kiến trúc của Yolov8 có thể được chia thành ba phần chính:

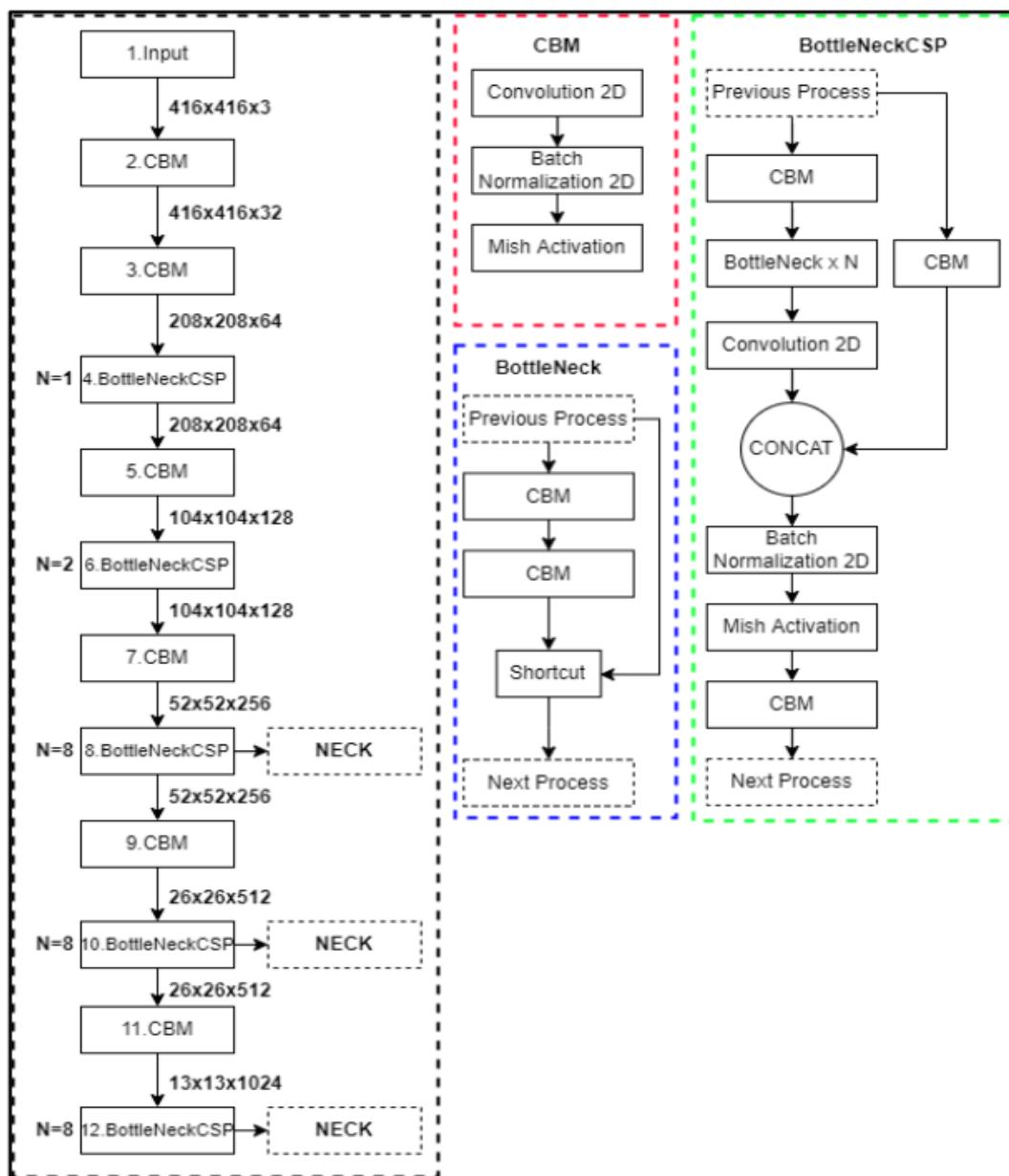
---

<sup>1</sup> <https://docs.ultralytics.com/models/yolov8/>

<sup>2</sup> <https://docs.ultralytics.com/>

## Backbone

Mạng backbone là nền tảng của mô hình Yolov8, chịu trách nhiệm trích xuất đặc trưng từ hình ảnh đầu vào. YOLOv8 sử dụng CSPDarknet53, một biến thể của Darknet, làm mạng backbone của nó. Kiến trúc CSPDarknet53 sử dụng chiến lược CSPNet (7) để phân chia bản đồ đặc trưng của lớp cơ bản thành hai phần và sau đó hợp nhất chúng thông qua một hệ thống phân cấp qua các giai đoạn. Việc sử dụng chiến lược phân chia và hợp nhất cho phép luồng gradient thông qua mạng mượt mà hơn. Hình 9 trình bày cấu trúc của CSPDarknet53, hình ảnh được trích từ bài nghiên cứu “Comparison of CSPDarknet53, CSPResNeXt-50, and EfficientNet-B0 Backbones on YOLO V4 as Object Detector (8)”.



Hình 9: Cấu trúc CSPDarkNet53

## Cấu trúc Neck

Yolov8 giới thiệu Path Aggregation Network (PANet) làm cấu trúc neck. PANet giúp thông tin trong hình ảnh di chuyển qua các độ phân giải không gian khác nhau một cách hiệu quả. Điều này có nghĩa là PANet giúp mô hình có khả năng nhận biết các đặc trưng ở nhiều tỉ lệ không gian khác nhau, từ đó cải thiện khả năng nhận diện đối tượng ở các kích thước và tỉ lệ khác nhau trong hình ảnh.

## Cấu trúc Head

Cấu trúc Head trong Yolov8 là phần cuối cùng của mạng, chịu trách nhiệm cho việc dự đoán các thông tin quan trọng liên quan đến các đối tượng trong hình ảnh. Cấu trúc này thường bao gồm nhiều “đầu phát hiện” (detection head). Mỗi detection head chịu trách nhiệm cho việc phát hiện các đối tượng ở một tỷ lệ khác nhau. Mỗi detection head thực hiện các công việc gồm: dự đoán bounding box, xác suất lớp và điểm số đối tượng (objectness score). Điểm số đối tượng đo lường khả năng một bounding box chứa một đối tượng thực sự. Điểm số này đánh giá mức độ tin cậy của mô hình về việc có hoặc không có đối tượng trong một bounding box cụ thể.

### 1.2.5. Yolov9

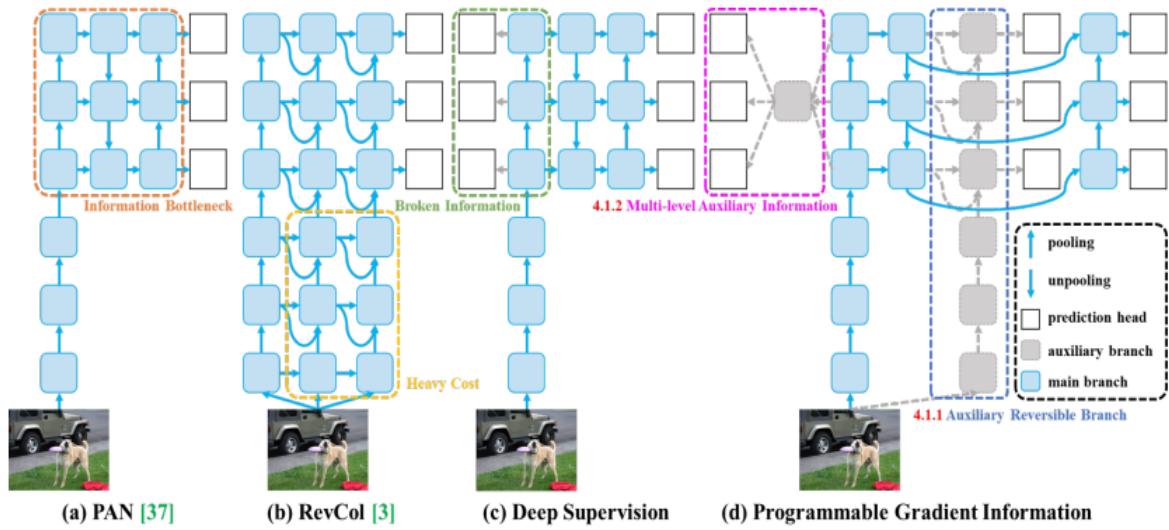
Yolov9<sup>3</sup> phiên bản mới nhất trong họ các mô hình của YOLO, được phát hành vào tháng 2/2024, do Chien-Yao Wang, I-Hau Yeh, Hong-Yuan Mark Liao phát triển. Đây là một mô hình phát hiện đối tượng thời gian thực được cải thiện nhằm vượt qua tất cả các phương pháp dựa trên convolution và transformer.

Để cải thiện độ chính xác, Yolov9 giới thiệu programmable gradient information (PGI) và Generalized Efficient Layer Aggregation Network (GELAN). PGI ngăn mất mát dữ liệu (data loss) và đảm bảo cập nhật gradient chính xác. GELAN tối ưu hóa các mô hình nhẹ với lập kế hoạch đường dốc gradient.

Để giải quyết vấn đề tắc nghẽn thông tin (data loss in the feed-forward process), Yolov9 đưa ra khái niệm mới là PGI. Mô hình tạo ra các gradient đáng tin cậy, thông qua một nhánh phụ có thể đảo ngược (auxiliary reversible branch). Các đặc trưng sâu (deep features) vẫn thực hiện nhiệm vụ mục tiêu và nhánh phụ tránh mất ý nghĩa do các đặc trưng đa đường (multi-path features). Hình 10 trình bày PGI và các kiến trúc mạng liên quan được trích từ bài báo “YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information” (9) – đây là bài báo chính thức của Yolov9 do nhóm tác giả phát hành.

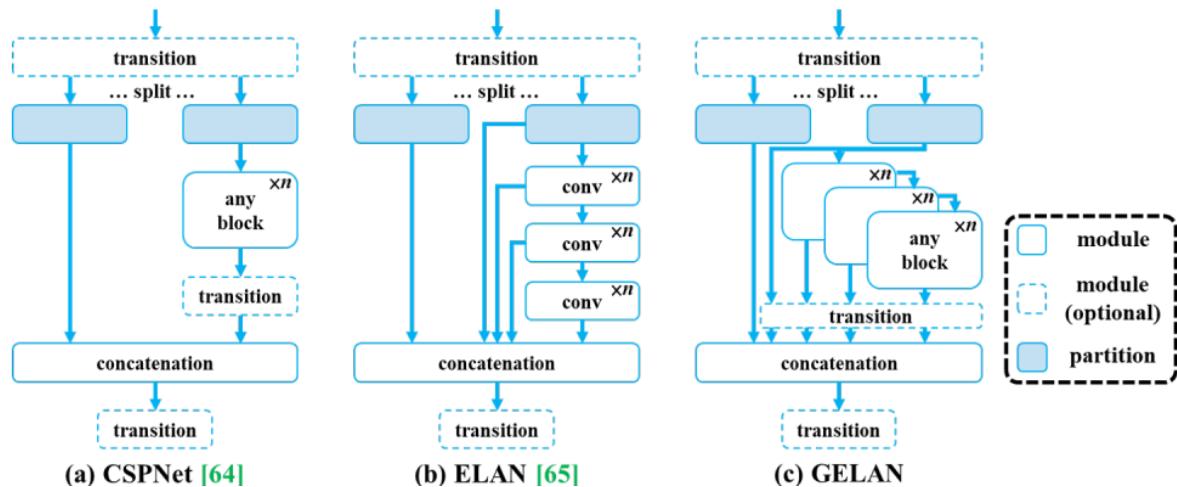
---

<sup>3</sup> <https://docs.ultralytics.com/models/yolov9/>



Hình 10: PGI và các kiến trúc mạng liên quan

Cơ chế PGI được đề xuất có thể áp dụng cho các mạng nơ-ron sâu của các kích thước khác nhau. Trong bài báo, nhóm tác giả đã thiết kế kiến trúc mạng GELAN – bằng cách kết hợp hai kiến trúc mạng nơ-ron: CSPNet (7) và ELAN. Đồng thời cân nhắc số lượng tham số, phức tạp tính toán, độ chính xác và tốc độ suy luận. Thiết kế cho phép người dùng chọn các khối tính toán phù hợp tùy ý cho các thiết bị suy luận khác nhau. Kiến trúc tổng thể của GELAN được trình bày trong hình 11, sơ đồ cũng được trích từ bài báo “YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information”.



Hình 11: Kiến trúc tổng thể của mạng GELAN

### 1.2.6. RT-DETR

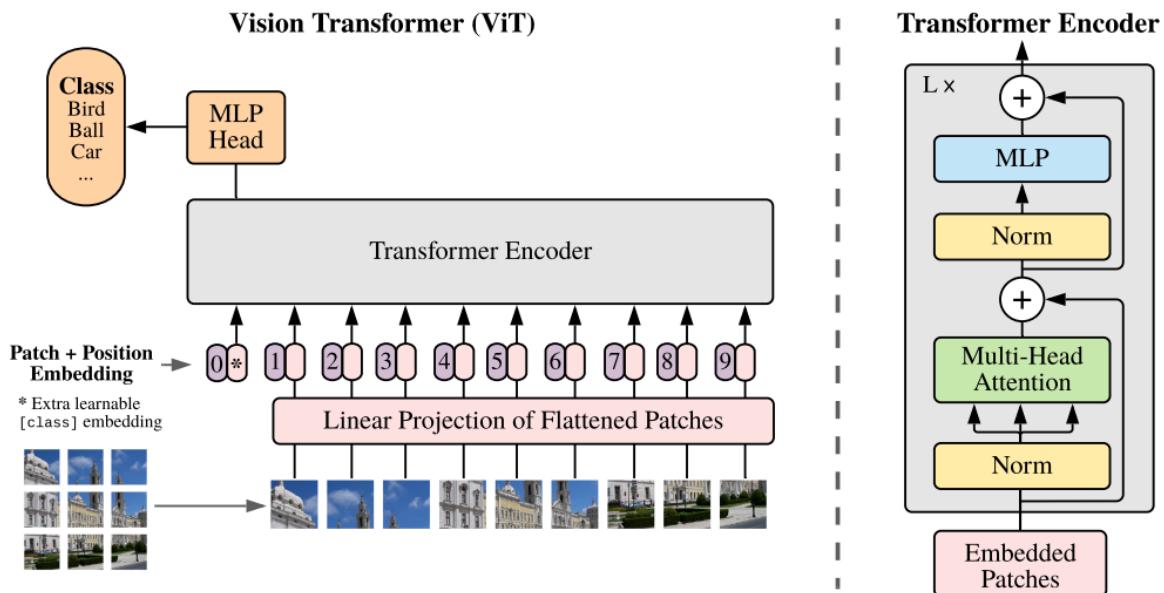
#### Vision Transformer (ViT)

Vào năm 2022, Vision Transformer (ViT) nổi lên như một giải pháp thay thế cạnh tranh so với các mạng CNN, vốn đang là ứng dụng tiên tiến trong thị giác máy tính. Các mô hình ViT được đánh giá là vượt trội hơn so với CNN gần 4 lần về hiệu quả tính toán và độ chính xác. Kiến trúc tổng thể của mô hình ViT được đưa ra như sau theo cách thức từng bước:

1. Chia hình ảnh thành các mảng (patch) với kích thước từng mảng cố định;
2. Làm phẳng các mảng hình ảnh;
3. Tạo các feature embedding có chiều thấp hơn từ các mảng hình ảnh phẳng này;
4. Bao gồm thứ tự các mảng;
5. Chuỗi feature embedding được làm đầu vào cho transformer encoder;
6. Thực hiện pre-train đối với mô hình ViT với các nhãn hình ảnh, sau đó được giám sát hoàn toàn trên một tập dữ liệu lớn;
7. Tinh chỉnh model trên bộ dữ liệu riêng của từng bài toán.

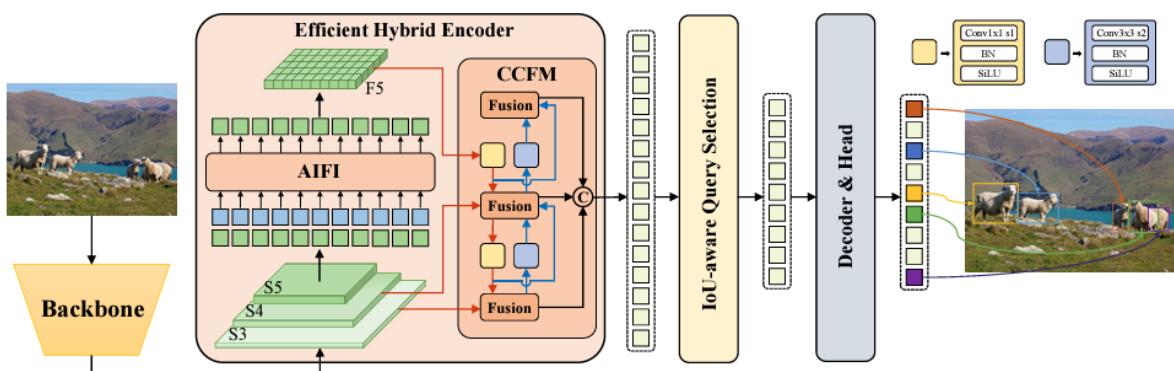
Hiệu suất của ViT phụ thuộc vào các quyết định như của trình tối ưu hóa, độ sâu mạng lưới và siêu tham số cụ thể của tập dữ liệu. Một mô hình ViT thông thường sử dụng tích chập 16x16 với stride 16. Trong khi đó, tích chập 3x3 với stride 2 làm tăng độ ổn định và nâng cao độ chính xác.

CNN chuyển đổi các điểm ảnh cơ bản thành một bản đồ đặc trưng. Sau đó, bản đồ đặc trưng được một trình mã hóa thành một chuỗi các tokens, rồi được nhập vào kiến trúc Transformer đó. Mô hình này sau đó áp dụng cơ chế chú ý để tạo ra một chuỗi các tokens đầu ra. Cuối cùng, 1 lớp projector kết nối các tokens đầu ra với bản đồ đặc trưng. Điều này vì vậy giúp làm giảm số lượng tokens cần được học, giảm chi phí đáng kể. Hình 12 mô tả tổng quan mô hình ViT, sơ đồ được trích từ bài báo “An image is worth 16x16 words: Transformers for image recognition at scale” (10).

*Hình 12: Tổng quan mô hình Vision Transformer*

### Real-Time Detection Transformer (RT-DETR)

RT-DETR (11) được phát triển bởi Baidu, là một bộ phát hiện đối tượng tiên tiến hoàn toàn mới mang lại hiệu suất thời gian thực cùng độ chính xác cao. Nó sử dụng công nghệ Vision Transformers (ViT) để xử lý hiệu quả các đặc trưng của đối tượng ở nhiều tỉ lệ khác nhau. RT-DETR không chỉ phân tách mà còn hòa nhập các mối quan hệ giữa các phần của đối tượng, giúp cải thiện khả năng nhận diện. Mô hình này có thể điều chỉnh tốc độ suy luận một cách linh hoạt mà không cần huấn luyện lại, phù hợp với nhiều mục đích sử dụng khác nhau. Với khả năng hoạt động ưu việt trên các nền tảng tăng tốc như CUDA với TensorRT, RT-DETR vượt trội hơn so với nhiều bộ phát hiện đối tượng thời gian thực khác. Hình 13 mô tả tổng quan kiến trúc của mô hình RT-DETR đăng bởi Ultralytics<sup>4</sup>.

*Hình 13: Tổng quan kiến trúc mô hình RT-DETR*

<sup>4</sup> <https://docs.ultralytics.com/models/rtdetr>

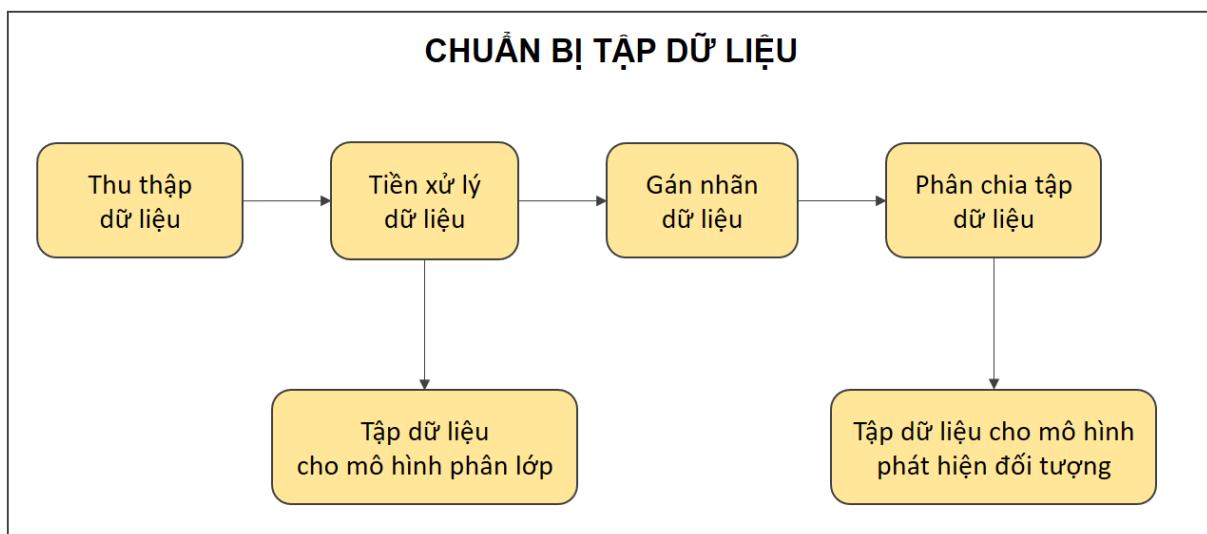
Với sơ đồ ở hình 13, ta thấy mô hình RT-DETR bao gồm ba phần chính: backbone, hybrid encoder, transformer decoder. Ba giai đoạn cuối cùng của mạng backbone {S3, S4, S5} được sử dụng làm đầu vào cho bộ mã hóa (hybrid encoder). Bộ mã hóa chuyển đổi đa tỉ lệ các đặc trưng thành một chuỗi các đặc trưng hình ảnh. Để thực hiện việc này, nó sử dụng hai mô-đun gồm: Attention based Intra Scale Feature Interaction (AIFI) và CNN based Cross-scale Feature-fusion Module (CCFM). Bộ mã hóa này thiết kế dựa trên ViT giúp giảm chi phí tính toán và cho phép phát hiện đối tượng theo thời gian thực. Tất cả đặc trưng hình ảnh từ bộ mã hóa không được dùng làm đầu vào cho bộ giải mã (transformer decoder). Thay vào đó, mô-đun IoU-aware Query Selection, chọn một tập hợp các đặc trưng hình ảnh làm đầu vào ban đầu cho bộ giải mã. Bộ giải mã, bao gồm các đầu dự đoán phụ, tạo ra các bounding box và độ tin cậy.

## CHƯƠNG 2

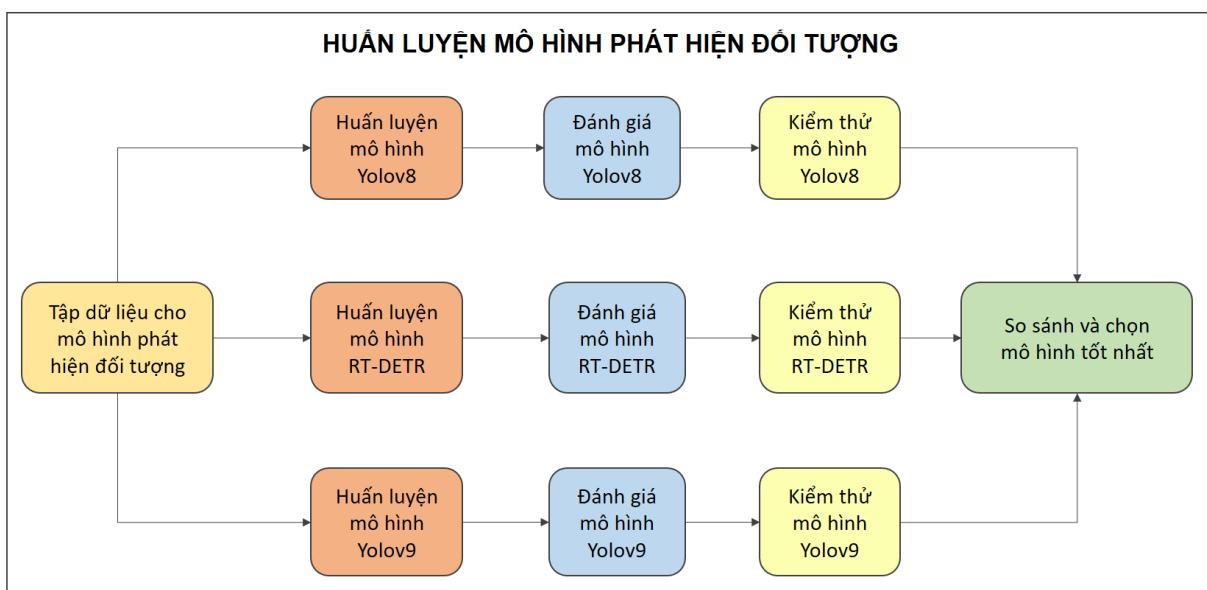
### THIẾT KẾ VÀ CÀI ĐẶT

#### 2.1. Thiết kế hệ thống

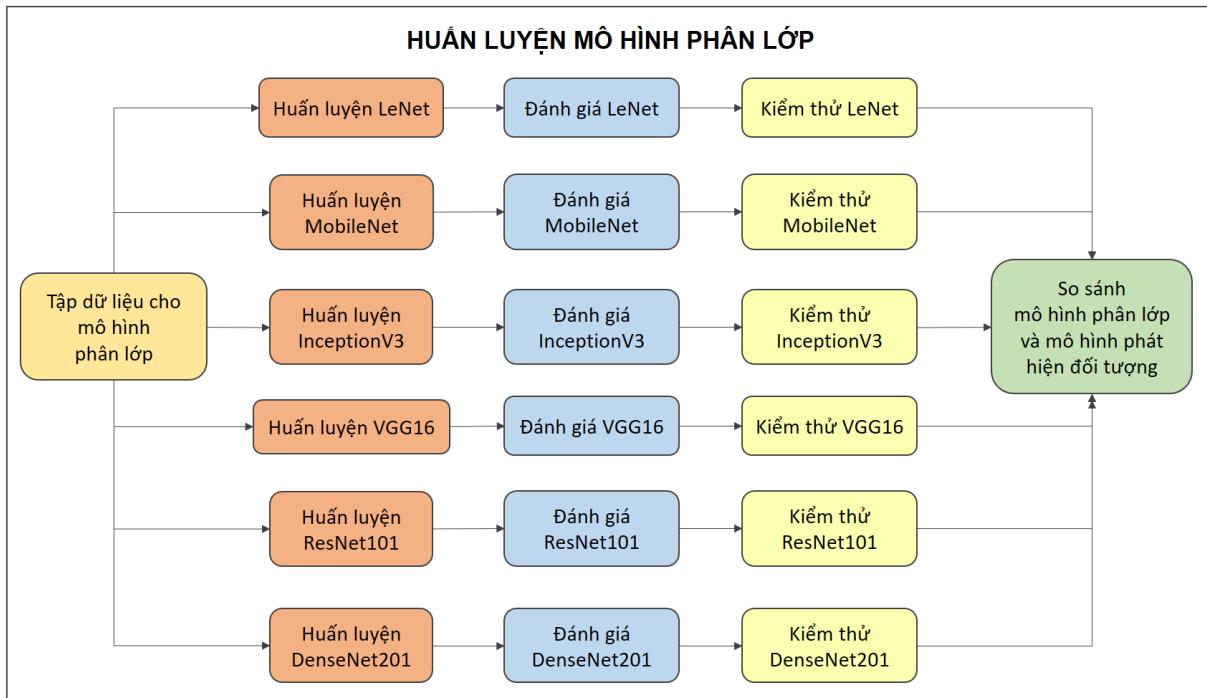
Đề tài của chúng tôi là “Xây dựng hệ thống xác định món ăn đặc sản Việt Nam dựa trên các thành phần nguyên liệu”. Thiết kế hệ thống này quá trình tổ chức và lập kế hoạch các bước thực hiện theo trình tự logic từ việc phân tích hình ảnh, nhận diện thành phần, đến việc xác định và phân loại món ăn. Mỗi bước được xác định và sắp xếp một cách cẩn thận để tạo thành một hệ thống hoàn chỉnh và hiệu quả, giúp đảm bảo tính chính xác và độ tin cậy của kết quả cuối cùng. Hình 14, 15, 16 lần lượt là các sơ đồ thể hiện quá trình chuẩn bị tập dữ liệu, huấn luyện mô hình phát hiện đối tượng, huấn luyện mô hình phân lớp.



Hình 14: Quá trình chuẩn bị tập dữ liệu



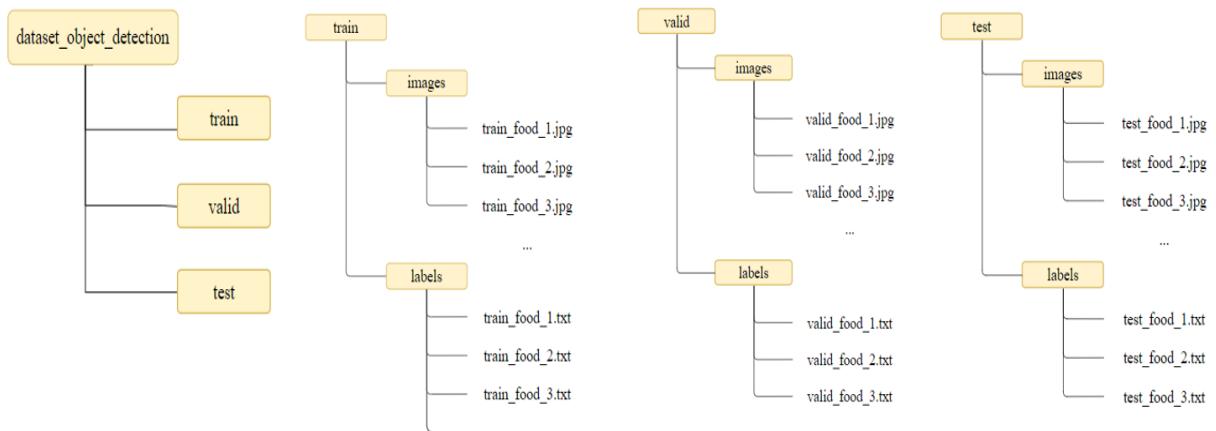
Hình 15: Quá trình huấn luyện mô hình phát hiện đối tượng



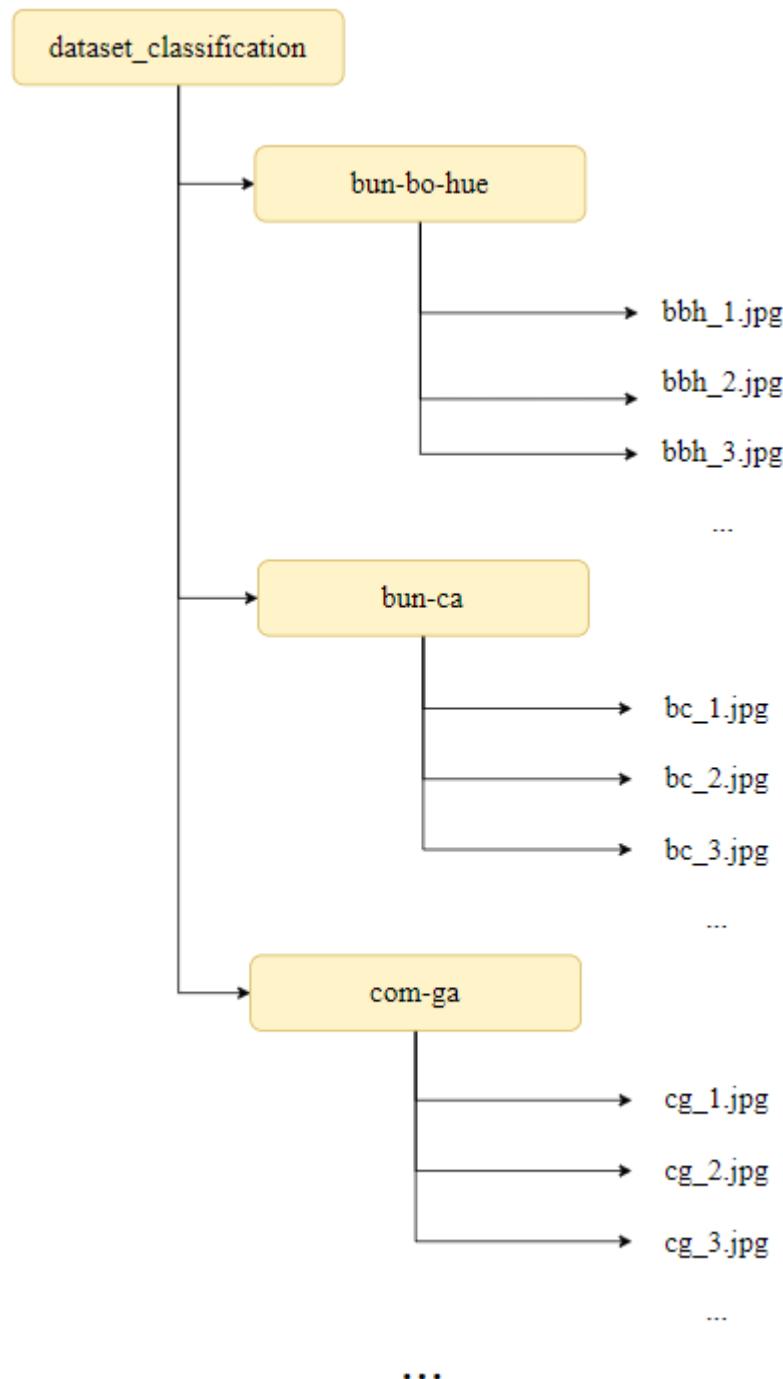
*Hình 16: Quá trình huấn luyện mô hình phân lớp với các kiến trúc: LeNet, MobileNet, InceptionV3, VGG16, ResNet101, DenseNet201*

## 2.2. Tập dữ liệu huấn luyện mô hình

Đề tài gồm hai tập dữ liệu về món ăn đặc sản Việt Nam. Một tập dữ liệu được sử dụng để huấn luyện mô hình nhận diện các thành phần nguyên liệu có trên món ăn. Tập dữ liệu còn lại phục vụ việc huấn luyện mô hình phân lớp CNN có khả năng phân loại món ăn đặc sản Việt Nam. Hình 17 mô tả cấu trúc cây thư mục của tập dữ liệu huấn luyện mô hình nhận diện thành phần nguyên liệu. Hình 18 mô tả cấu trúc cây thư mục của tập dữ liệu huấn luyện mô hình CNN phân loại món ăn đặc sản Việt Nam.

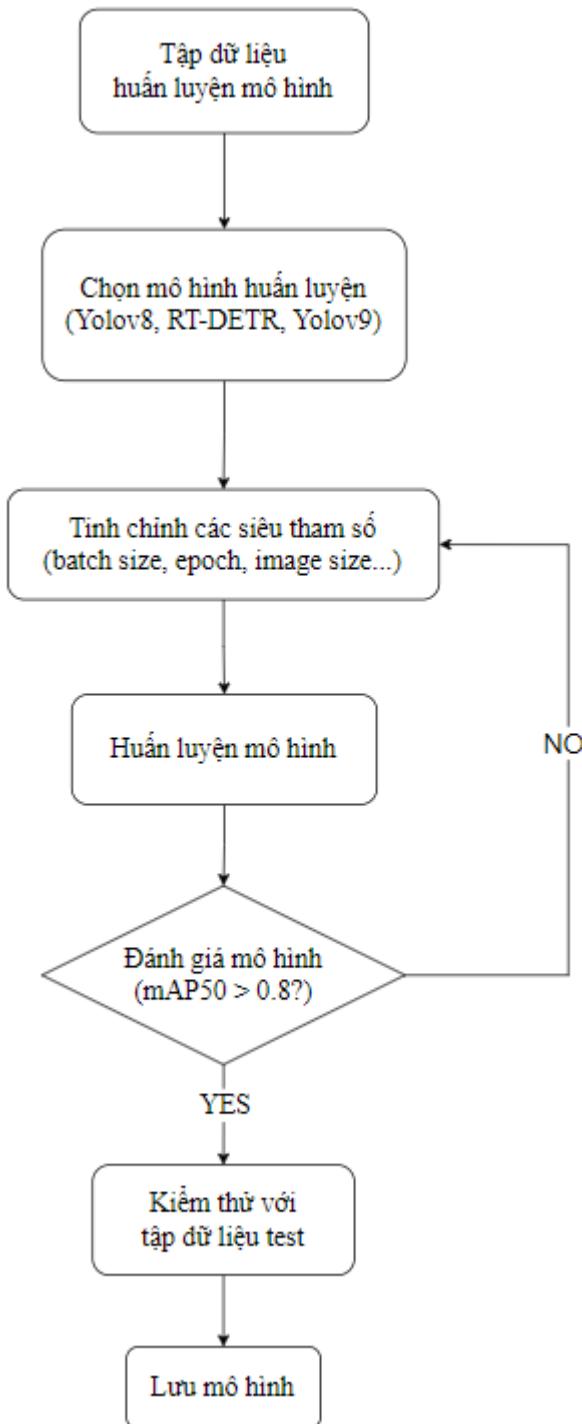


*Hình 17: Cấu trúc cây thư mục tập dữ liệu huấn luyện mô hình phát hiện nguyên liệu*



Hình 18: Cấu trúc cây thư mục tập dữ liệu huấn luyện mô hình phân lớp  
món ăn đặc sản Việt Nam

### 2.3. Xây dựng mô hình huấn luyện nhận diện thành phần nguyên liệu

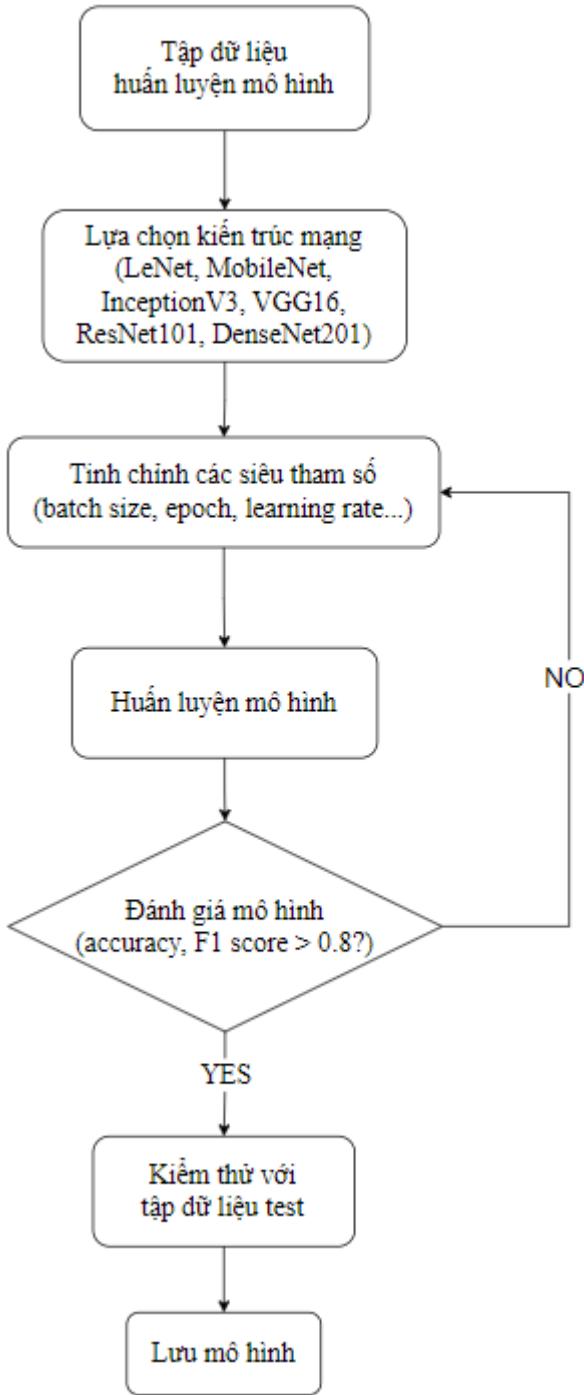


Hình 19: Xây dựng mô hình nhận diện các thành phần nguyên liệu

Sơ đồ ở hình 19 cho thấy tập dữ liệu được sử dụng để huấn luyện ba mô hình phát hiện đối tượng gồm: Yolov8, RT-DETR, Yolov9. Các mô hình được tinh chỉnh với nhiều siêu tham số khác nhau như: batch size, số lần học, kích cỡ ảnh đầu vào... Các mô hình còn được fine tuning từ các mô hình đã được huấn luyện trước, giúp tận dụng được kiến thức đã được mô hình học từ dữ liệu lớn, từ đó giảm thiểu thời gian và tài nguyên cần thiết cho việc huấn luyện mạng nơ-ron, hiệu suất của mô hình cũng được tăng.

Chỉ số mAP50 (12) được sử dụng làm tiêu chí đánh giá hiệu suất, độ chính xác của mô hình. Trường hợp mô hình có chỉ số mAP50 trên 0.8 thì có thể đưa mô hình vào nhận diện các thành phần nguyên liệu. Nếu mô hình chưa đạt yêu cầu thì tinh chỉnh lại các siêu tham số cho phù hợp và tiếp tục huấn luyện mô hình. Kiểm thử mô hình với tập dữ liệu test. Nếu kết quả kiểm thử có độ chính xác cao thì lưu mô hình lại để tiếp tục sử dụng.

## 2.4. Xây dựng mô hình huấn luyện phân lớp món ăn đặc sản



Hình 20: Xây dựng mô hình phân lớp món ăn đặc sản

Sơ đồ ở hình 20 cho thấy tập dữ liệu được sử dụng để huấn luyện sáu mô hình phân lớp CNN gồm các kiến trúc mạng nơ-ron: LeNet (13), MobileNet (14), InceptionV3 (15), VGG16 (16), ResNet101 (17), DenseNet201 (18). Các mô hình được tinh chỉnh với nhiều siêu tham số khác nhau như: batch size, số lần học, tốc độ học... Lý do huấn luyện các mô hình phân lớp để so sánh giữa phương pháp chính của đề tài – “Xây dựng hệ thống xác định món ăn đặc sản Việt Nam dựa trên các thành phần nguyên liệu” và phương pháp phân lớp.

Xây dựng mô hình phân lớp bằng các thuật toán học sâu CNN, đang là xu hướng được sử dụng rộng rãi như là hệ quả của sự phát triển các kiến trúc học sâu. Đề tài nghiên cứu của chúng tôi huấn luyện thêm các mô hình phân lớp CNN để so sánh với mô hình chính của nghiên cứu. Từ đây có thể thấy được phương pháp xử lý nào sẽ mang lại kết quả nổi trội hơn.

## 2.5. Xây dựng công thức tính khoảng cách

Bảng 2: Bộ luật các thành phần nguyên liệu có trong các món ăn đặc sản Việt Nam

| i  | Món ăn             | Nguyên liệu ưu tiên | Nguyên liệu chính                           | Nguyên liệu phụ            | Tỉnh thành         |
|----|--------------------|---------------------|---|----------------------------|--------------------|
| 1  | Bún cá             | bún                 | cá  | tôm, rau răm               | An Giang           |
| 2  | Hủ tiếu Mỹ Tho     | hủ tiếu             | thịt băm, thịt heo, gan                     | trứng, tôm                 | Tiền Giang         |
| 3  | Bún nước lèo       | bún                 | cá, tôm, thịt heo quay                      | rau răm                    | Sóc Trăng          |
| 4  | Cơm tấm Long Xuyên | cơm                 | sườn, bì, trứng                             | dưa chua                   | An Giang           |
| 5  | Bún hải sản bè bè  | bún                 | tôm tí                                      | tôm, mực                   | Vũng Tàu           |
| 6  | Bánh hỏi heo quay  | bánh hỏi            | thịt heo quay                               |                            | Cần Thơ            |
| 7  | Cơm gà             | cơm                 | thịt gà                                     | rau răm                    | Quảng Nam          |
| 8  | Cao lầu            | mì, thịt heo        | bánh đa                                     |                            | Quảng Nam          |
| 9  | Mì Quảng           | mì                  | bánh đa, trứng                              | thịt gà, thịt heo, cá, tôm | Quảng Nam, Đà Nẵng |
| 10 | Bún bò Huế         | bún                 | thịt bò                                     |                            | Thừa Thiên Huế     |
| 11 | Phở Hà Nội         | phở                 | thịt bò                                     |                            | Hà Nội             |
| 12 | Bún mực            | bún                 | mực   | tôm                        | Phú Yên            |
| 13 | Bún mọc            | bún                 | viên mọc                                    | thịt heo                   | Hà Nội             |
| 14 | Bún đậu mắm tôm    | bún                 | chả cốt, dồi sụn, đậu hũ, thịt heo, chả giò |                            | Hà Nội             |

Bảng 2 mô tả lại bộ luật gồm các thành phần nguyên liệu tương ứng với từng món ăn đặc sản Việt Nam đã được trình bày ở chương 1. Danh sách thành phần nguyên liệu của món ăn được chia làm ba loại theo thứ tự có tính chất quyết định đến món ăn gồm: nguyên liệu ưu tiên, nguyên liệu chính, nguyên liệu phụ.

**Nguyên liệu ưu tiên** là nguyên liệu có tính chất quyết định nhất đến tên món ăn. Ví dụ ở các món được chế biến từ bún, sẽ có nguyên liệu ưu tiên là bún, ở các món được chế biến từ mì sẽ có nguyên liệu ưu tiên là mì... Nguyên liệu ưu tiên giúp phân biệt rạch ròi các món bún, mì, phở, cơm... với nhau.

**Nguyên liệu chính** là các thành phần nguyên liệu cũng có tính chất quyết định đến tên món ăn, nhưng thấp hơn nguyên liệu ưu tiên. Ví dụ ở món bún cá sẽ có nguyên liệu chính là cá, ở món bún mực sẽ có thành phần nguyên liệu chính là mực... Nguyên liệu chính giúp phân biệt rõ ràng hơn các món ăn nếu chúng được chế biến cùng thành phần nguyên liệu ưu tiên. Điển hình là trường hợp các món được chế biến từ bún, có bún cá, bún nước lèo, bún mực...

**Nguyên liệu phụ** là các thành phần nguyên liệu thường được ăn kèm trong món ăn đó, cùng với nguyên liệu ưu tiên và nguyên liệu chính. Ví dụ ở món bún cá và cơm gà, thực khách thường ăn kèm với rau răm, khi dùng món bún mực, thực khách cũng thường được ăn kèm với tôm. Các thành phần nguyên liệu phụ này cũng có tính chất quyết định đến món ăn, đóng vai trò nhận biết món ăn.

Bài toán của chúng tôi là sử dụng mô hình AI để nhận diện các thành phần nguyên liệu có trên ảnh món ăn, từ đó thu được một tập hợp là kết quả chứa các thành phần nguyên liệu mà AI phát hiện được.

Gọi kết quả tập hợp chứa các thành phần nguyên liệu được mô hình phát hiện là DE. Gọi món ăn trong ảnh đầu vào cần xác định là “**món ăn X**”. Gọi các món ăn được liệt kê trong bộ luật là  $F_i$ . Với  $i = 1$   $F_1$  là món bún cá,  $i = 2$   $F_2$  là món hủ tiếu Mỹ Tho,  $i = 3$   $F_3$  là món bún nước lèo,  $i = 4$   $F_4$  là món cơm tấm Long Xuyên... Mỗi món ăn sẽ gồm có ba tập hợp lần lượt là  $P_i$ ,  $M_i$ ,  $E_i$ . Trong đó,  $P_i$  là tập hợp chứa các nguyên liệu ưu tiên,  $M_i$  là tập hợp chứa các nguyên liệu chính,  $E_i$  là tập hợp chứa các nguyên liệu phụ.

Tên gọi cụ thể của món ăn X là món gì sẽ được xác định dựa trên việc đo khoảng cách giữa món ăn X này, đến  $F_i$  được liệt kê có trong bộ luật ở bảng 2. Công thức tính khoảng cách được chúng tôi thiết kế như sau:

$$P_i^- = P_i \setminus DE \quad d^-(X, F_i) = (|P_i^-| \times 10) + |M_i^-| - (|E_i^+| \times 0.5)$$

$$M_i^- = M_i \setminus DE \quad d^+(X, F_i) = |R_i|$$

$$E_i^+ = E_i \cap DE \quad d_i(X, F_i) = d^-(X, F_i) + d^+(X, F_i)$$

$$R_i = DE \setminus P_i \setminus M_i \setminus E_i$$

Chú thích:

X: Món ăn cần xác định;

DE: Tập hợp chứa các nguyên liệu của X mà mô hình phát hiện được;

$F_i$ : Món ăn thứ i được liệt kê trong bộ luật,  $i = 1..14$ ;

$P_i$ : Tập hợp chứa các nguyên liệu ưu tiên của  $F_i$ ;

$M_i$ : Tập hợp chứa các nguyên liệu chính của  $F_i$ ;

$E_i$ : Tập hợp chứa các nguyên liệu phụ của  $F_i$ ;

$P_i^-$ : Tập hợp chứa các nguyên liệu có trong  $P_i$  nhưng không có trong DE;

$M_i^-$ : Tập hợp chứa các nguyên liệu có trong  $M_i$  nhưng không có trong DE;

$E_i^+$ : Tập hợp chứa các nguyên liệu cùng có trong  $E_i$  và DE;

$R_i$ : Tập hợp các nguyên liệu có trong DE nhưng không có trong  $P_i$ ,  $M_i$ ,  $E_i$ ;

$d_i(X, F_i)$ : Khoảng cách từ X đến  $F_i$ .

## 2.6. Cài đặt hệ thống

Cài đặt hệ thống là quá trình tiến hành lập trình, đáp ứng mục tiêu của nghiên cứu này là xây dựng một mô hình hệ thống thông minh có khả năng xác định món ăn đặc sản của Việt Nam dựa trên các thành phần nguyên liệu. Quá trình cài đặt hệ thống sẽ đi qua các công việc: thu thập dữ liệu, tiền xử lý dữ liệu, phân chia tập dữ liệu phục vụ giai đoạn huấn luyện mô hình, huấn luyện mô hình, cài đặt giải thuật tính khoảng cách đã được thiết kế ở trên, triển khai mô hình thành ứng dụng.

### 2.6.1. Thu thập dữ liệu

Dữ liệu hình ảnh của nghiên cứu này tập trung vào hình ảnh của 14 món ăn đặc sản Việt Nam gồm: bún cá, hủ tiếu Mỹ Tho, bún nước lèo, cơm tấm Long Xuyên, bún hải sản bè bè, bánh hỏi heo quay, cơm gà, cao lầu, mì Quảng, bún bò Huế, phở Hà Nội, bún mực, bún mọc, bún đậu mắm tôm. Ngoài ra còn thu thập thêm hình ảnh của một số món ăn khác. Nguồn ảnh được thu thập từ Google Images, kết hợp với chụp ảnh món ăn ngoài thực tế bằng nhiều thiết bị chụp ảnh khác nhau – nhằm tạo nên sự đa dạng cho tập dữ liệu.

### 2.6.2. Tiền xử lý dữ liệu

Ảnh sau khi thu thập sẽ trải qua quá trình tiền xử lý dữ liệu bao gồm: cắt ảnh, tăng cường dữ liệu (xoay ảnh), gán nhãn dữ liệu. Ảnh gốc có chứa các phần thừa không liên quan đến nội dung món ăn, cần cắt ảnh để tập trung vào món ăn trong ảnh. Một số lượng ảnh cũng được xử lý thay đổi bằng cách xoay ảnh với nhiều góc độ khác nhau. Cuối cùng ảnh được gán nhãn các thành phần nguyên liệu bằng LabelImg, chuẩn bị tập dữ liệu phục vụ huấn luyện mô hình phát hiện đối tượng. Hình 21 mô tả quá trình tiền xử lý của một mẫu dữ liệu. Tổng số lượng nhãn (lớp) mà mô hình được huấn luyện là 27 nhãn lần lượt là các nguyên liệu sau: cá, bún, hủ tiếu, tôm, thịt heo, gan, thịt heo quay, thịt bò, trứng, thịt bầm, tôm tít, mực, viên mọc, bánh hỏi, dưa chua, phở, cơm, sườn, bì, thịt gà, rau răm, mì, bánh đa, chả cốt, đồi sụn, đậu hũ, chả giò.



Hình 21: Tiền xử lý và gán nhãn dữ liệu

### 2.6.3. Phân chia tập dữ liệu

Một phần tập dữ liệu sẽ được sử dụng làm tập kiểm thử (testing). Số còn lại sẽ được sử dụng phục vụ giai đoạn huấn luyện mô hình. Tập huấn luyện (training) và tập xác thực (validation) sẽ được phân bổ theo tỷ lệ 70% - 30%. Bảng 3 thống kê số lượng mẫu dữ liệu có trong từng tập huấn luyện, xác thực, kiểm thử.

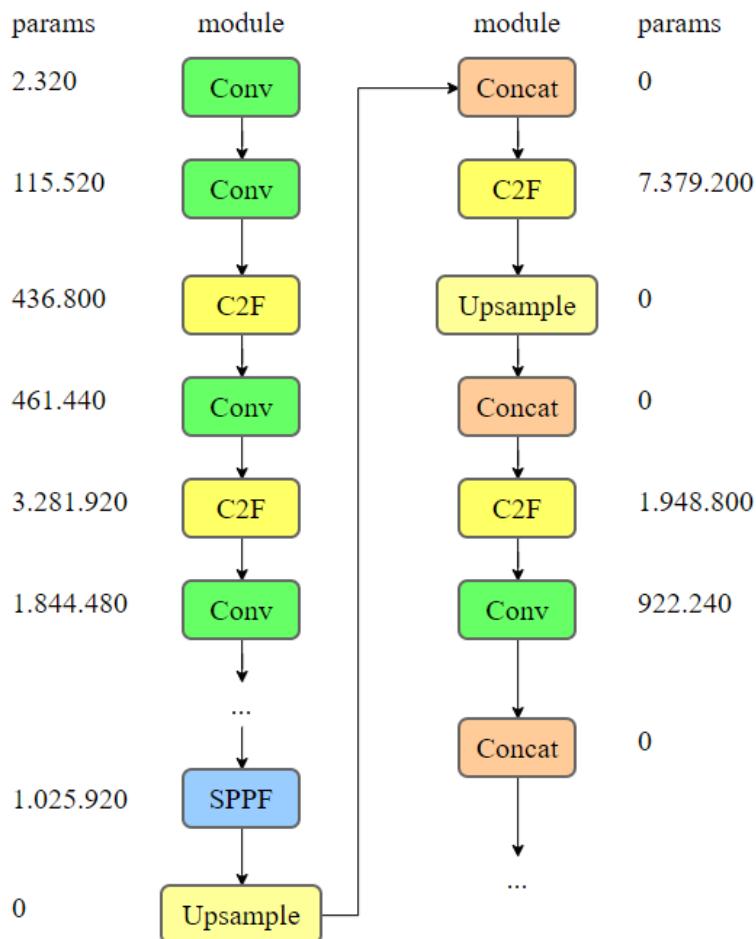
Bảng 3: Thống kê số lượng mẫu trong tập dữ liệu

| Tập dữ liệu        | Số lượng mẫu |
|--------------------|--------------|
| Huấn luyện (train) | 2494         |
| Xác thực (valid)   | 864          |
| Kiểm thử (test)    | 81           |

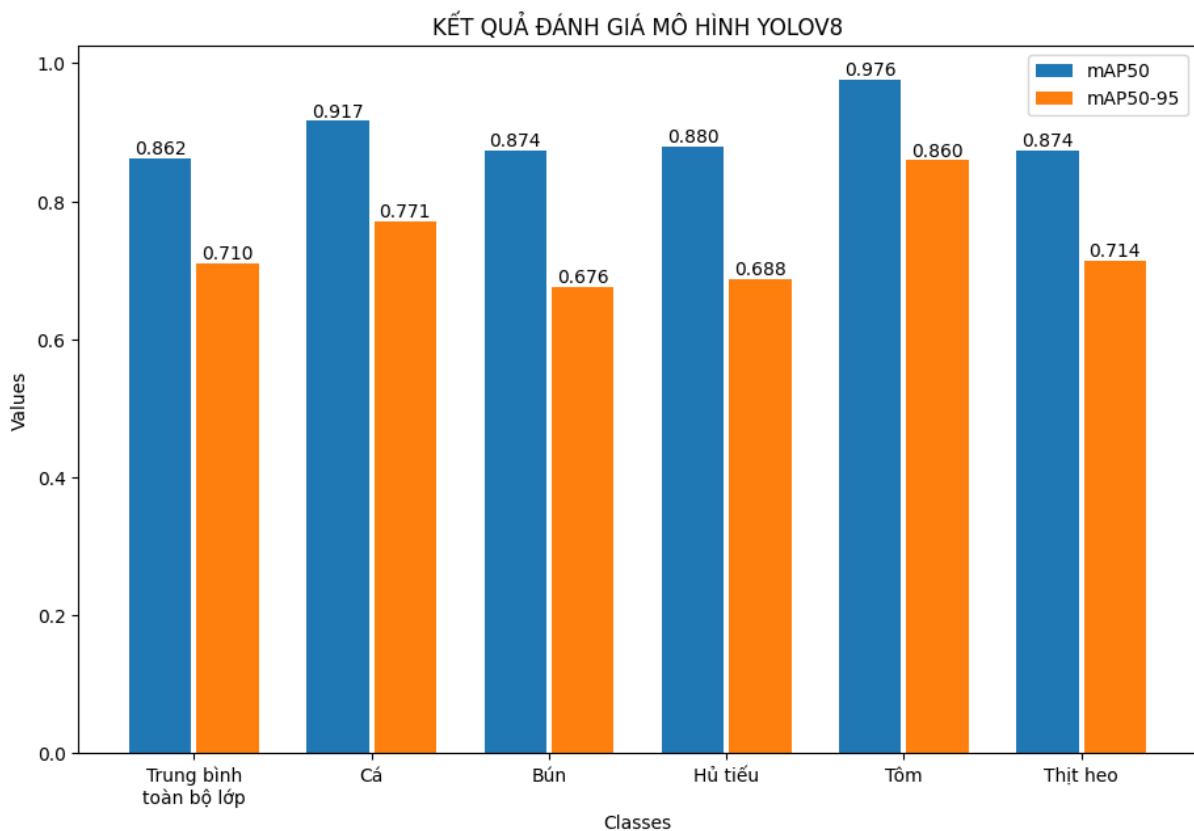
## 2.6.4. Huấn luyện mô hình

### 2.6.4.1. Yolov8

Thực hiện huấn luyện mô hình Yolov8, fine tuning từ mô hình “yolov8x.pt” đã được huấn luyện trước. Sau khi huấn luyện thu được mô hình có **268 tầng, 68.149.569 tham số**. Do số lượng tầng mạng nơ-ron quá lớn, chúng tôi trình bày một số tầng trong cấu trúc mạng của mô hình Yolov8 ở hình 22. Đồng thời với số lượng nhãn của bài toán lớn (27 nguyên liệu), chúng tôi cũng chỉ trình bày kết quả đánh giá mô hình sau khi huấn luyện trên một số lớp sau: trung bình toàn bộ các lớp nguyên liệu, lớp cá, lớp bún, lớp hủ tiếu, lớp tôm. Hình 23 trình bày kết quả đánh giá mô hình.



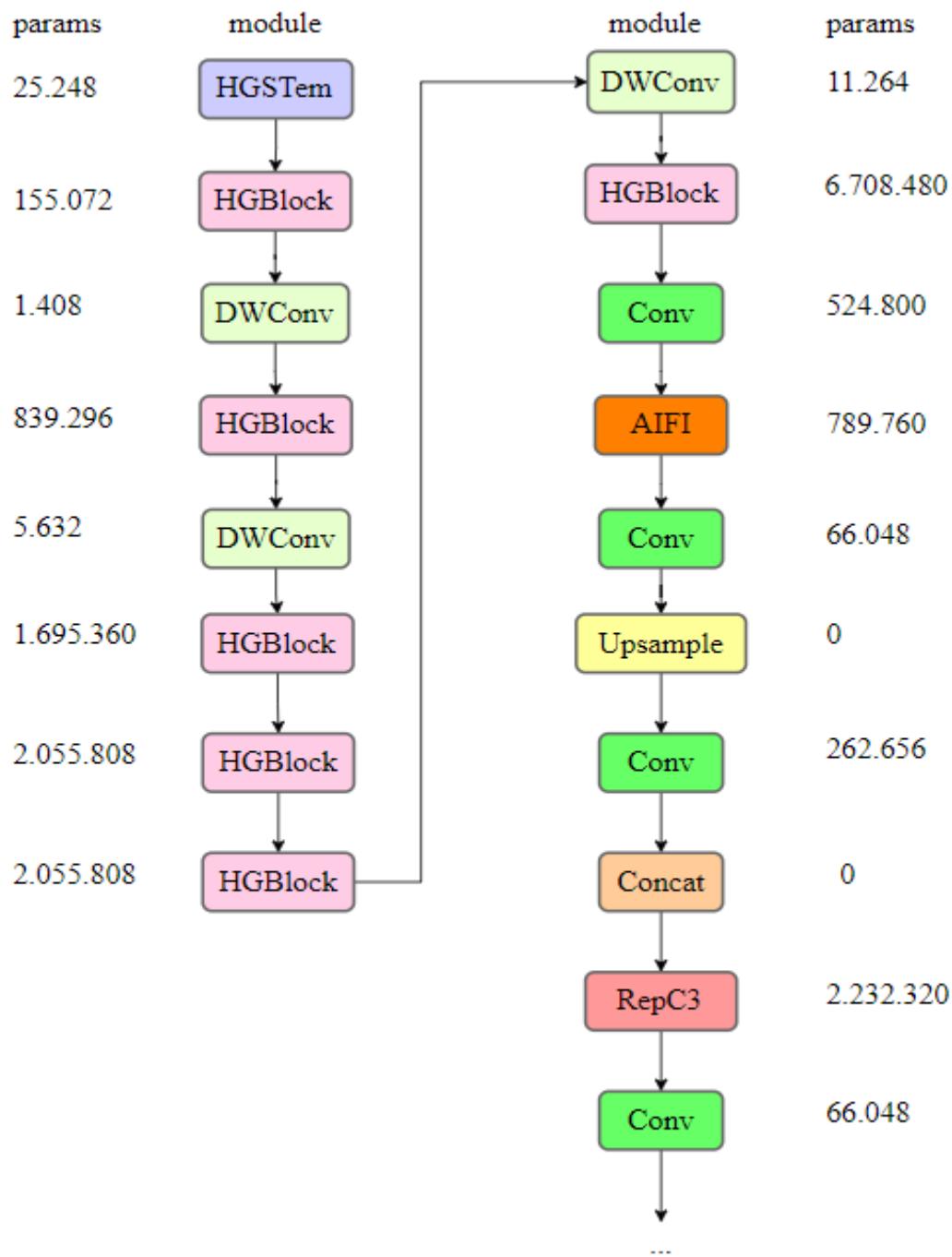
Hình 22: Cấu trúc mạng nơ-ron của mô hình Yolov8



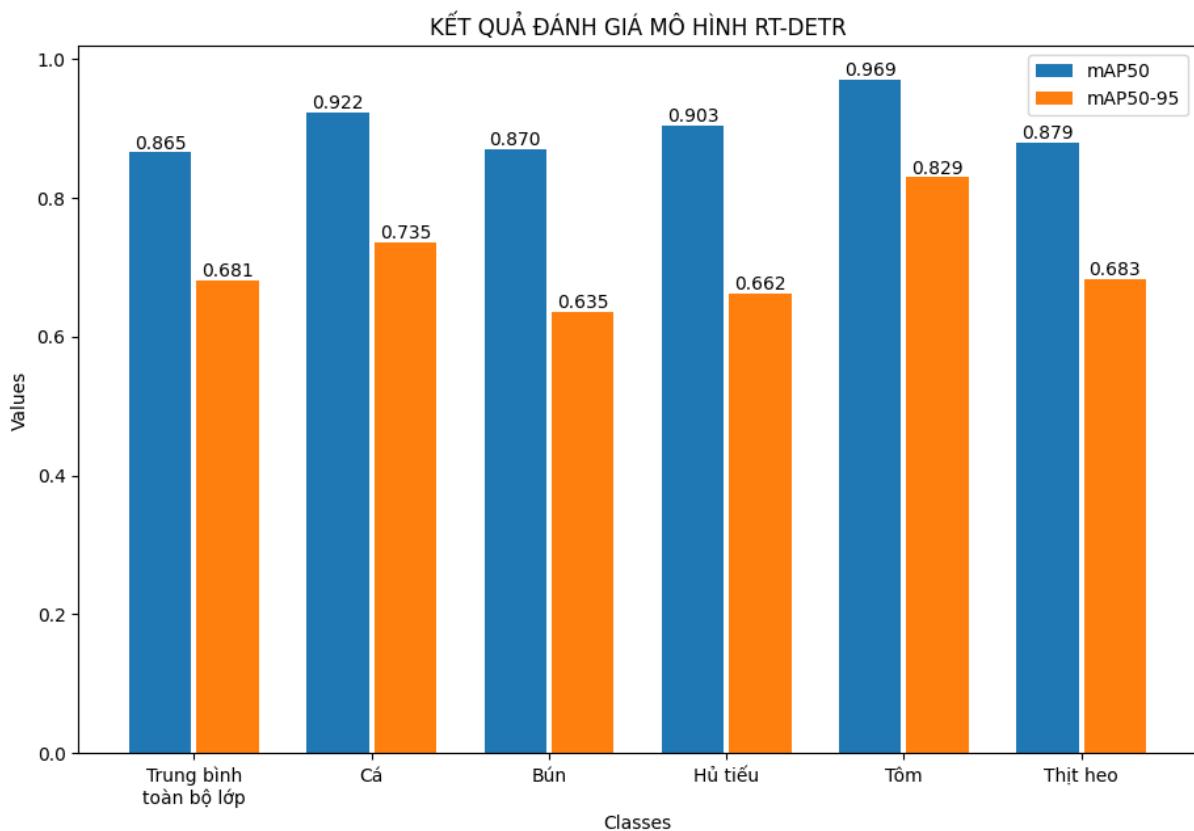
*Hình 23: Kết quả đánh giá mô hình Yolov8 sau khi huấn luyện*

#### 2.6.4.2. RT-DETR

Thực hiện huấn luyện mô hình RT-DETR, fine tuning từ mô hình “rtdetr-l.pt” đã được huấn luyện trước. Sau khi huấn luyện thu được mô hình có **498 tầng, 32.039.225 tham số**. Do số lượng tầng mạng nơ-ron quá lớn, chúng tôi trình bày một số tầng trong cấu trúc mạng của mô hình RT-DETR ở hình 24. Đồng thời với số lượng nhãn của bài toán lớn (27 nguyên liệu), chúng tôi cũng chỉ trình bày kết quả đánh giá mô hình sau khi huấn luyện trên một số lớp sau: trung bình toàn bộ các lớp nguyên liệu, lớp cá, lớp bún, lớp hủ tiêu, lớp tôm. Hình 25 trình bày kết quả đánh giá mô hình.



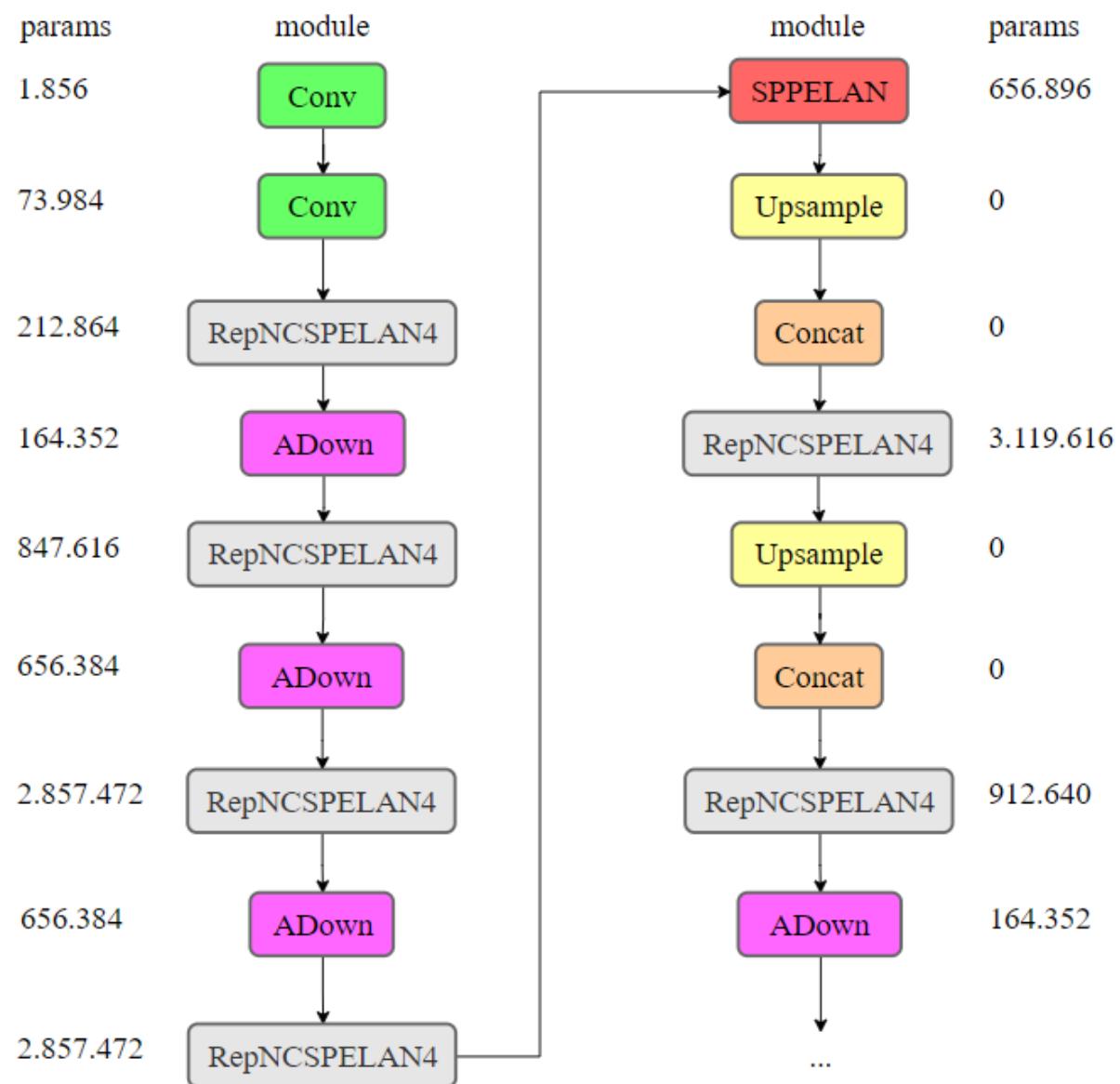
Hình 24: Cấu trúc mạng nơ-ron của mô hình RT-DETR



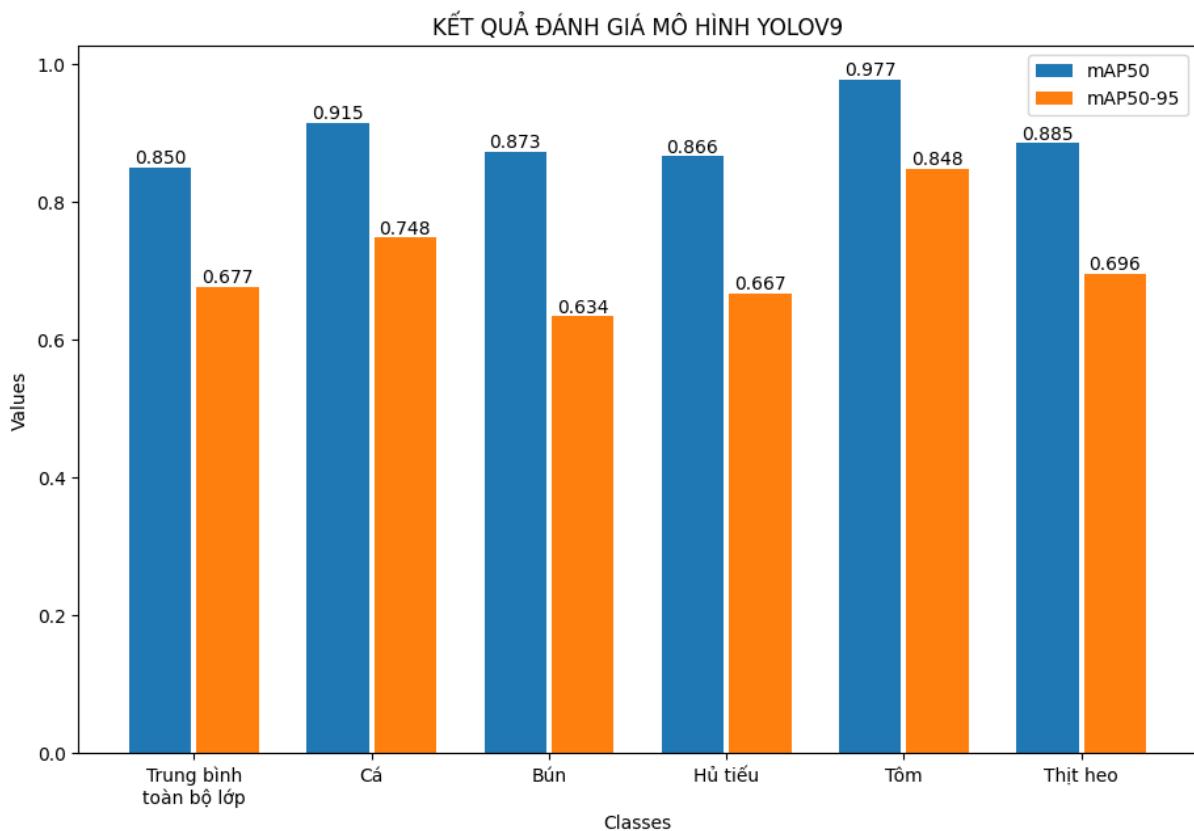
*Hình 25: Kết quả đánh giá mô hình RT-DETR sau khi huấn luyện*

#### 2.6.4.3. Yolov9

Thực hiện huấn luyện mô hình Yolov9, fine tuning từ mô hình “yolov9c.pt” đã được huấn luyện trước. Sau khi huấn luyện thu được mô hình có **384 tầng, 25.340.065 tham số**. Do số lượng tầng mạng nơ-ron quá lớn, chúng tôi trình bày một số tầng trong cấu trúc mạng của mô hình Yolov9 ở hình 26. Đồng thời với số lượng nhãn của bài toán lớn (27 nguyên liệu), chúng tôi cũng chỉ trình bày kết quả đánh giá mô hình sau khi huấn luyện trên một số lớp sau: trung bình toàn bộ các lớp nguyên liệu, lớp cá, lớp bún, lớp hủ tiêu, lớp tôm. Hình 27 trình bày kết quả đánh giá mô hình.



Hình 26: Cấu trúc mạng của mô hình Yolov9



*Hình 27: Kết quả đánh giá mô hình Yolov9 sau khi huấn luyện*

#### 2.6.4.4. Các mô hình phân lớp

Thực hiện huấn luyện mô hình phân lớp CNN, phân loại món ăn đặc sản, với nhiều kiến trúc học sâu khác nhau gồm: LeNet, MobileNet, InceptionV3, VGG16, ResNet101, DenseNet201. Các mô hình cũng được tinh chỉnh với nhiều siêu tham số khác nhau như: số lần học, batch size, tốc độ học... Bảng 4, 5, 6, 7, 8, 9 trình bày cấu trúc mạng của lần lượt các mô hình: LeNet, MobileNet, InceptionV3, VGG16, ResNet101, DenseNet201. Bảng 10 thể hiện độ chính xác accuracy, F1 score của các mô hình phân lớp.

*Bảng 4: Cấu trúc mạng của LeNet*

| <b>Layer</b>         | <b>Output Shape</b>  | <b>Param</b> |
|----------------------|----------------------|--------------|
| Conv2D               | (None, 256, 256, 32) | 2.432        |
| MaxPooling2D         | (None, 128, 128, 32) | 0            |
| Conv2D               | (None, 124, 124, 48) | 38.448       |
| MaxPooling2D         | (None, 62, 62, 48)   | 0            |
| Flatten              | (None, 184.512)      | 0            |
| Dense                | (None, 256)          | 47.235.328   |
| Dense                | (None, 84)           | 21.588       |
| Dense                | (None, 14)           | 1.190        |
| Tổng tham số         |                      | 47.298.986   |
| Trainable params     |                      | 47.298.986   |
| Non-trainable params |                      | 0            |

*Bảng 5: Cấu trúc mạng của MobileNet*

| <b>Layer</b>           | <b>Output Shape</b> | <b>Param</b> |
|------------------------|---------------------|--------------|
| mobilenet              | (None, 8, 8, 1024)  | 3.228.864    |
| GlobalAveragePooling2D | (None, 1024)        | 0            |
| Dropout                | (None, 1024)        | 0            |
| Flatten                | (None, 1024)        | 0            |
| Dense                  | (None, 14)          | 14.350       |
| Tổng tham số           |                     | 3.243.214    |
| Trainable params       |                     | 14.350       |
| Non-trainable params   |                     | 3.228.864    |

*Bảng 6: Cấu trúc mạng của InceptionV3*

| <b>Layer</b>           | <b>Output Shape</b> | <b>Param</b> |
|------------------------|---------------------|--------------|
| inceptionv3            | (None, 6, 6, 2048)  | 21.802.784   |
| GlobalAveragePooling2D | (None, 2048)        | 0            |
| Dropout                | (None, 2048)        | 0            |
| Flatten                | (None, 2048)        | 0            |
| Dense                  | (None, 14)          | 28.686       |
| Tổng tham số           |                     | 21.831.470   |
| Trainable params       |                     | 28.686       |
| Non-trainable params   |                     | 21.802.784   |

*Bảng 7: Cấu trúc mạng của VGG16*

| <b>Layer</b>           | <b>Output Shape</b> | <b>Param</b> |
|------------------------|---------------------|--------------|
| vgg16                  | (None, 8, 8, 512)   | 14.714.688   |
| GlobalAveragePooling2D | (None, 512)         | 0            |
| Dropout                | (None, 512)         | 0            |
| Flatten                | (None, 512)         | 0            |
| Dense                  | (None, 14)          | 7.182        |
| Tổng tham số           |                     | 14.721.870   |
| Trainable params       |                     | 7.182        |
| Non-trainable params   |                     | 14.714.688   |

*Bảng 8: Cấu trúc mạng của ResNet101*

| <b>Layer</b>           | <b>Output Shape</b> | <b>Param</b> |
|------------------------|---------------------|--------------|
| resnet101              | (None, 8, 8, 2048)  | 42.658.176   |
| GlobalAveragePooling2D | (None, 2048)        | 0            |
| Dropout                | (None, 2048)        | 0            |
| Flatten                | (None, 2048)        | 0            |
| Dense                  | (None, 14)          | 28.686       |
| Tổng tham số           |                     | 42.686.862   |
| Trainable params       |                     | 28.686       |
| Non-trainable params   |                     | 42.658.176   |

*Bảng 9: Cấu trúc mạng của DenseNet201*

| <b>Layer</b>           | <b>Output Shape</b> | <b>Param</b> |
|------------------------|---------------------|--------------|
| densenet201            | (None, 8, 8, 1920)  | 18.321.984   |
| GlobalAveragePooling2D | (None, 1920)        | 0            |
| Dropout                | (None, 1920)        | 0            |
| Flatten                | (None, 1920)        | 0            |
| Dense                  | (None, 14)          | 26.894       |
| Tổng tham số           |                     | 18.348.878   |
| Trainable params       |                     | 26.894       |
| Non-trainable params   |                     | 18.321.984   |

*Bảng 10: Độ chính xác mô hình phân loại món ăn đặc sản Việt Nam*

| Mô hình     | Accuracy (%) | Precision (%) | Recall (%)   | F1 score (%) |
|-------------|--------------|---------------|--------------|--------------|
| LeNet       | 25.79        | 25.57         | 25.79        | 25.12        |
| MobileNet   | <b>76.84</b> | <b>79.08</b>  | <b>76.84</b> | <b>76.42</b> |
| InceptionV3 | 65.96        | 66.94         | 65.96        | 65.65        |
| VGG16       | 60.88        | 60.29         | 60.88        | 58.15        |
| ResNet101   | 19.3         | 23.02         | 19.3         | 14.03        |
| DenseNet201 | <b>78.6</b>  | <b>79.57</b>  | <b>78.6</b>  | <b>78.43</b> |

### 2.6.5. Cài đặt giải thuật tính khoảng cách

```

Function find_foodname(set_DE):
    distances = []
    for row in df_rules:
        set_P = set(row[“Priorities”])
        set_M = set(row[“Ingredients”])
        set_E = set(row[“Secondary Ingredients”])
        P_negative = set_P \ set_DE
        M_negative = set_M \ set_DE
        E_positive = set_E ∩ set_DE
        combine = set_P ∪ set_M ∪ set_E
        set_R = set_DE \ combine
        d_negative = (len(P_negative) * 10) + len(M_negative)
                    – (len(E_positive) * 0.5)
        d_positive = len(set_R)
        distance = d_negative + d_positive
        distances.append(distance)

    min_d = min(distances)
    min_indices = []
    for i, value in enumerate(distances):
        if value == min_d:
            min_indices.append(i)
    predicted_food = []
    for index in min_indices:
        predicted_food.append(df_rules.loc[index, “Food”])
    return predicted_food

```

Trong đó:

set\_DE: tập hợp các thành phần nguyên liệu mà mô hình phát hiện được;

distances: danh sách chứa tất cả khoảng cách từ món ăn cần xác định đến các món ăn được liệt kê trong bộ luật;

df\_rules: bộ luật gồm tên món ăn và các thành phần nguyên liệu của món ăn;

set\_P: tập hợp chứa các nguyên liệu ưu tiên;

set\_M: tập hợp chứa các nguyên liệu chính;

set\_E: tập hợp chứa các nguyên liệu phụ;

P\_negative: tập hợp chứa các nguyên liệu có trong set\_P mà không có trong set\_DE;

M\_negative: tập hợp chứa các nguyên liệu có trong set\_M mà không có trong set\_DE;

E\_positive: tập hợp chứa các nguyên liệu đều thuộc cả set\_E và set\_DE;

set\_R: tập hợp các nguyên liệu có trong set\_DE nhưng không có trong hợp của set\_P, set\_M, set\_E;

distance: khoảng cách được tính từ món ăn cần xác định đến món ăn đang xét hiện tại trong bộ luật;

min\_d: giá trị khoảng cách ngắn nhất trong distances;

min\_indices: danh sách chứa chỉ số (index) của các giá trị ngắn nhất trong distances;

predicted\_food: danh sách chứa kết quả là tên các món ăn được xác định.

### **2.6.6. Triển khai mô hình**

Chúng tôi quyết định đưa mô hình vào sử dụng thực tế bằng cách triển khai mô hình thành ứng dụng web. Backend được xây dựng bằng Flask, frontend được xây dựng bằng VueJS, với sự hỗ trợ tạo giao diện người dùng của thư viện Vuetify.

Flask là một microframework web nhẹ được viết bằng Python. Nó được biết đến với sự đơn giản, linh hoạt và dễ sử dụng. Với sự đơn giản, khả năng mở rộng và cộng đồng lớn, Flask là một framework mạnh mẽ có thể đáp ứng nhu cầu của nhiều dự án khác nhau.

VueJS là một framework JavaScript mã nguồn mở được sử dụng để xây dựng giao diện người dùng (UI) tương tác. Nó được biết đến với sự đơn giản, dễ học và dễ sử dụng, khiến nó trở thành lựa chọn phổ biến cho các nhà phát triển web ở mọi cấp độ kinh nghiệm.

Vuetify là một thư viện giao diện người dùng (UI) mã nguồn mở được xây dựng cho VueJS. Nó cung cấp một bộ sưu tập phong phú các thành phần giao diện người dùng được thiết kế sẵn, đẹp mắt và dễ sử dụng, giúp bạn nhanh chóng tạo ra các ứng dụng web hiện đại và đáp ứng.

## CHƯƠNG 3

### KẾT QUẢ, ĐÁNH GIÁ VÀ GIAO DIỆN

#### **3.1. Môi trường thực nghiệm**

Môi trường thực nghiệm là một phần quan trọng trong việc nghiên cứu và phát triển dự án. Đây là nơi mà các nhà nghiên cứu có thể thiết lập và thử nghiệm các thuật toán, mô hình, và phương pháp mới trên dữ liệu thực tế hoặc dữ liệu mô phỏng để đánh giá hiệu suất của chúng. Để đáp xây dựng và kiểm tra hệ thống và mô hình huấn luyện, chúng tôi đã thực hiện cài đặt các thư viện (thư viện được thống kê trình bày trong bảng 11) và chạy thực nghiệm trên máy tính có cấu hình như sau:

##### **3.1.1. Cấu hình máy**

1. Google Colab: 4 vCPU Intel Xeon E2-8276 (2.50 GHz), 16 GB RAM, 1x Tesla T4 GPU (16 GB HBM);
2. Laptop Dell Inspiron N5502: CPU Intel Core i7 1165G7, RAM 8GB, ổ cứng 512 GB SSD, hệ điều hành window 10.

##### **3.1.2. Các thư viện sử dụng**

Bảng 11: Các thư viện được sử dụng

| Tên thư viện  | Phiên bản |
|---------------|-----------|
| ultralytics   | 8.2.11    |
| opencv-python | 4.8.0.76  |
| numpy         | 1.25.2    |
| pandas        | 2.0.3     |
| matplotlib    | 3.7.1     |
| scikit-learn  | 1.2.2     |
| keras         | 2.15.0    |
| PIL           | 9.4.0     |

### 3.2. Kết quả kiểm tra, đánh giá

#### 3.2.1. Kết quả kiểm thử mô hình phát hiện thành phần nguyên liệu

Để lựa chọn mô hình phát hiện nguyên liệu, chúng tôi đã thực hiện chạy mô hình với nhiều siêu tham số khác nhau. Ở đây, hệ thống đã kiểm tra với ba mô hình phát hiện nguyên liệu, cụ thể bao gồm: Yolov8, RT-DETR, Yolov9. Bảng 12 thống kê độ chính xác (mAP50, mAP50 – 95), dung lượng của ba mô hình.

*Bảng 12: Thống kê độ chính xác, dung lượng của ba mô hình*

*phát hiện thành phần nguyên liệu*

| Mô hình | mAP50 | mAP50 – 95 | Dung lượng (MB) |
|---------|-------|------------|-----------------|
| Yolov8  | 0.862 | 0.71       | 130             |
| RT-DETR | 0.865 | 0.681      | 63.1            |
| Yolov9  | 0.85  | 0.677      | 49.2            |

Chúng Tôi thực hiện kiểm thử các mô hình này bằng cách cho các mô hình xác định tên món ăn trong tập dữ liệu ảnh các món ăn đặc sản. Bảng 13 thống kê dự đoán của ba mô hình phát hiện thành phần nguyên liệu.

Bảng 13: Thống kê kiểm thử các mô hình phát hiện thành phần nguyên liệu

| Món ăn             | Ảnh dự đoán | Yolov8                  |                    | Yolov9                  |                    | RT-DETR                 |                    |
|--------------------|-------------|-------------------------|--------------------|-------------------------|--------------------|-------------------------|--------------------|
|                    |             | Dự đoán đúng            | Tỷ lệ dự đoán đúng | Dự đoán đúng            | Tỷ lệ dự đoán đúng | Dự đoán đúng            | Tỷ lệ dự đoán đúng |
| Bún cá             | 111         | 110                     | 99.1%              | 111                     | 100%               | 110                     | 99.10%             |
| Hủ tiếu Mỹ Tho     | 60          | 50                      | 83.33%             | 51                      | 85%                | 53                      | 88.33%             |
| Bún nước lèo       | 120         | 106                     | 88.33%             | 105                     | 87.50%             | 105                     | 87.50%             |
| Cơm tấm Long Xuyên | 195         | 195                     | 100%               | 195                     | 100%               | 194                     | 99.49%             |
| Bún hải sản bè bè  | 174         | 168                     | 96.55%             | 167                     | 95.98%             | 167                     | 95.98%             |
| Bánh hỏi heo quay  | 96          | 96                      | 100%               | 96                      | 100%               | 96                      | 100%               |
| Cơm gà             | 172         | 172                     | 100%               | 172                     | 100%               | 172                     | 100%               |
| Cao lầu            | 165         | 154                     | 93.33%             | 154                     | 93.33%             | 155                     | 93.94%             |
| Mì Quảng           | 150         | 141                     | 94%                | 139                     | 92.67%             | 143                     | 95.33%             |
| Bún bò Huế         | 144         | 143                     | 99.31%             | 144                     | 100%               | 143                     | 99.31%             |
| Phở Hà Nội         | 177         | 177                     | 100%               | 176                     | 99.44%             | 177                     | 100%               |
| Bún mực            | 84          | 83                      | 98.81%             | 83                      | 98.81%             | 82                      | 97.62%             |
| Bún mọc            | 104         | 103                     | 99.04%             | 104                     | 100%               | 103                     | 99.04%             |
| Bún đậu mắm tôm    | 146         | 139                     | 95.21%             | 140                     | 95.89%             | 138                     | 94.52%             |
|                    |             | Độ chính xác trung bình | 96.79%             | Độ chính xác trung bình | 96.79%             | Độ chính xác trung bình | 96.84%             |

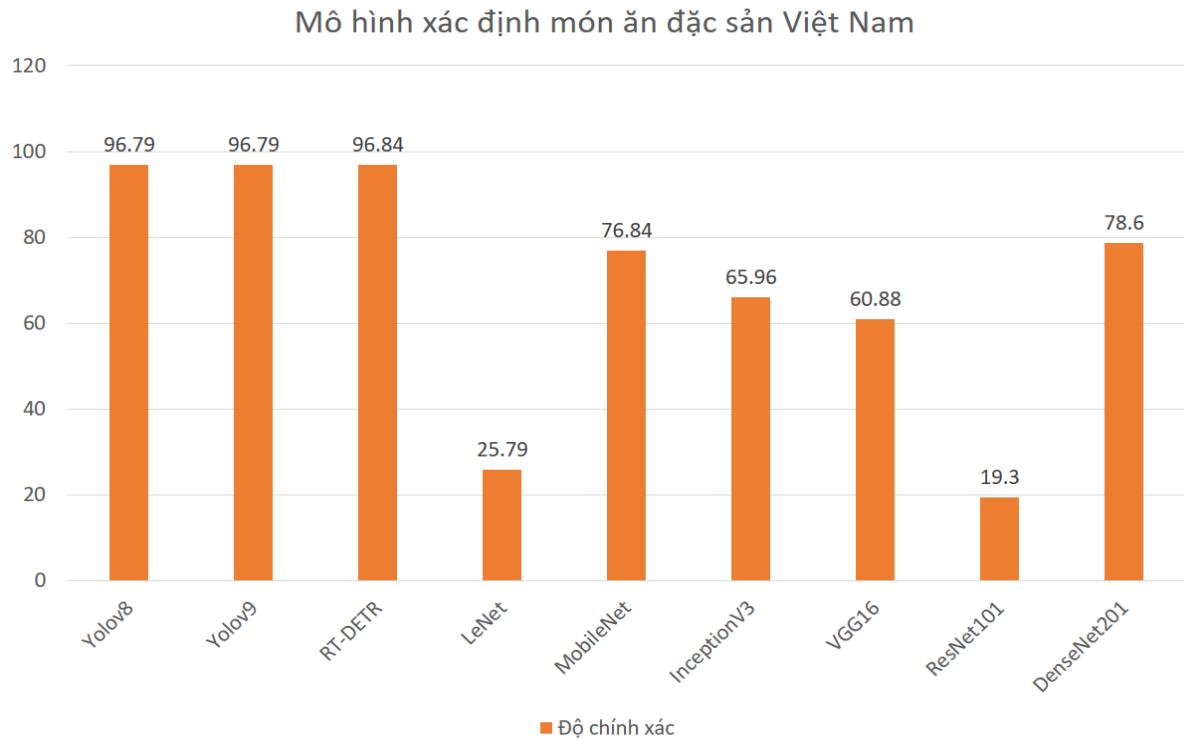
### 3.2.2. Kết quả kiểm thử mô hình phân lớp món ăn đặc sản

Mô hình phân lớp món ăn đặc sản được huấn luyện để so sánh hiệu suất với mô hình phát hiện nguyên liệu. Tập dữ liệu phân lớp được chia theo tỷ lệ 7/3. 70% dữ liệu được sử dụng để huấn luyện mô hình, 30% dữ liệu được sử dụng để kiểm thử – đánh giá độ chính xác của mô hình. Bảng 14 thống kê lại kết quả kiểm thử độ chính xác của các mô hình phân lớp.

Bảng 14: Độ chính xác mô hình phân loại món ăn đặc sản Việt Nam

| Mô hình     | Accuracy (%) | Precision (%) | Recall (%)   | F1 score (%) |
|-------------|--------------|---------------|--------------|--------------|
| LeNet       | 25.79        | 25.57         | 25.79        | 25.12        |
| MobileNet   | <b>76.84</b> | <b>79.08</b>  | <b>76.84</b> | <b>76.42</b> |
| InceptionV3 | 65.96        | 66.94         | 65.96        | 65.65        |
| VGG16       | 60.88        | 60.29         | 60.88        | 58.15        |
| ResNet101   | 19.3         | 23.02         | 19.3         | 14.03        |
| DenseNet201 | <b>78.6</b>  | <b>79.57</b>  | <b>78.6</b>  | <b>78.43</b> |

Hiệu suất của mô hình phân lớp CNN đạt dưới 80% thấp hơn mô hình phát hiện đối tượng. Với kết quả này, chúng tôi kết luận được phương pháp chính của nghiên cứu này: **xác định món ăn đặc sản dựa trên các thành phần nguyên liệu, mang lại kết quả vượt trội hơn phương pháp phân lớp CNN** – vốn đang được sử dụng rộng rãi. Biểu đồ so sánh độ chính xác giữa tất cả các mô hình được thể hiện trong hình 28.



Hình 28: So sánh các mô hình xác định món ăn đặc sản Việt Nam

### 3.3. Giao diện

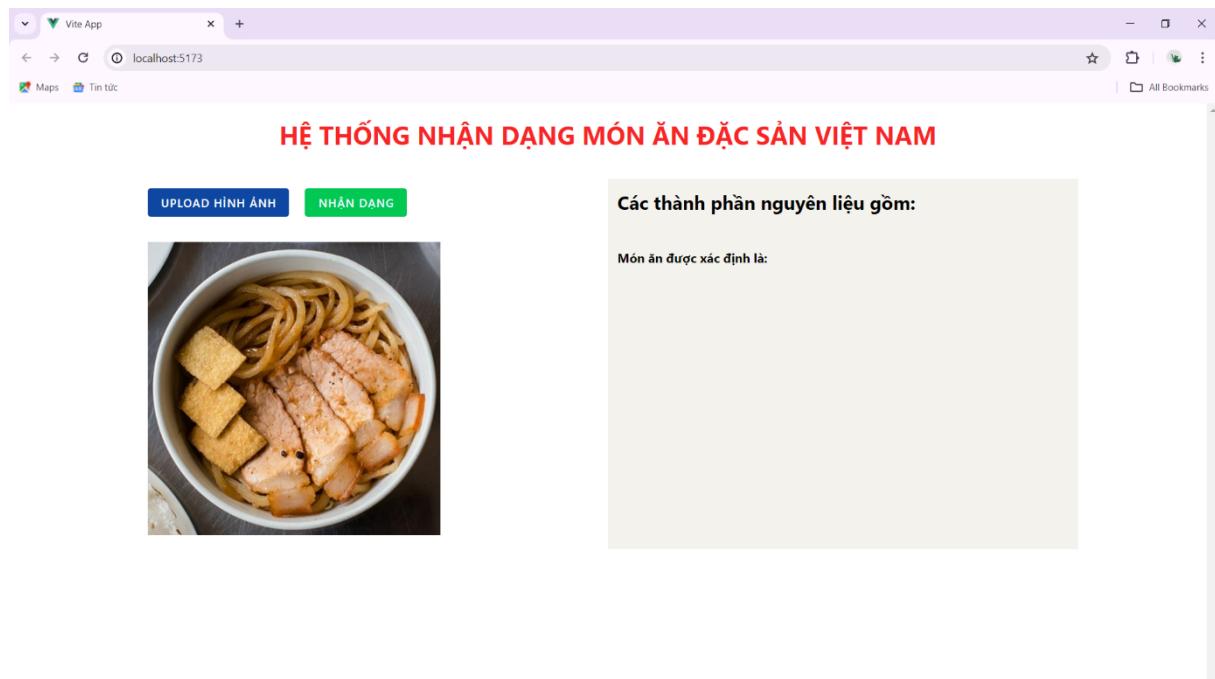
Giao diện là một phần không thể thiếu đối với hệ thống, dù hệ thống có ít tính năng thì vẫn có giao diện. Giao diện nhằm giúp cho người dùng tương tác với hệ thống thuận tiện và dễ dàng hơn.



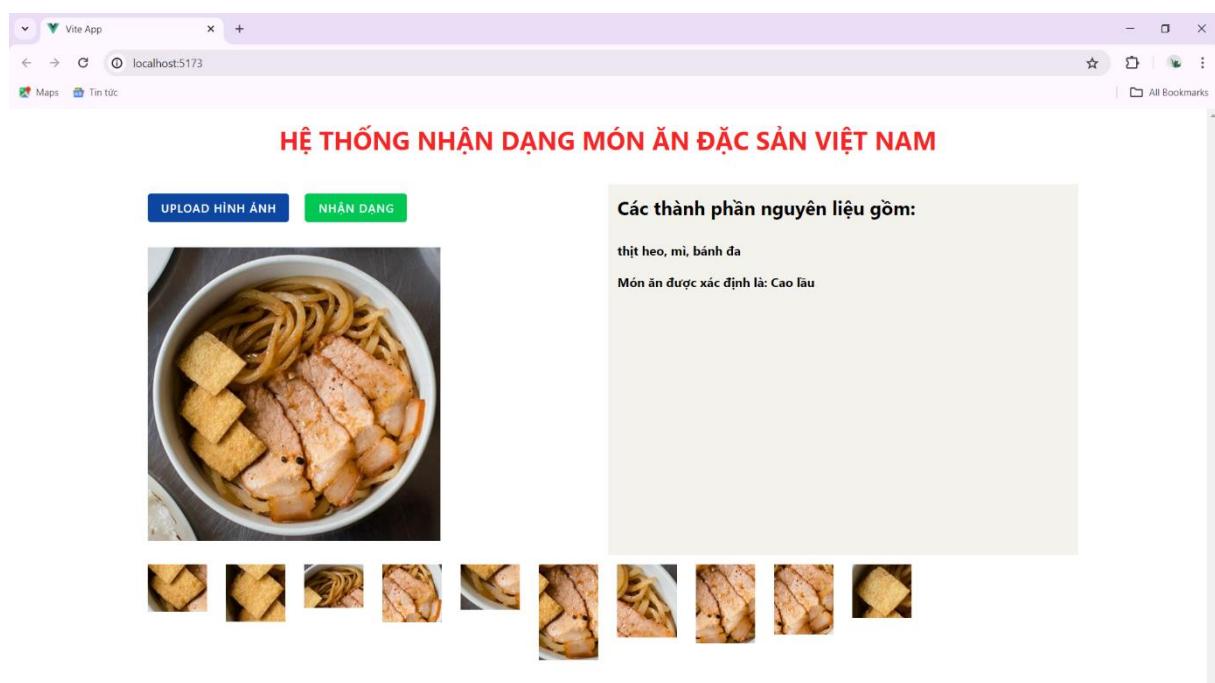
Hình 29: Giao diện hệ thống

Qua hình 29, chúng ta thấy được giao diện hệ thống có các thành phần sau:

- Nút “UPLOAD HÌNH ẢNH” cho phép người dùng upload hình ảnh món ăn lên hệ thống. Kiểm thử hoạt động sau khi nhấn nút “UPLOAD AN IMAGE” được trình bày ở hình 30;
- Nút “NHẬN DẠNG” có tác dụng kích hoạt hệ thống, cho mô hình AI nhận dạng các thành phần nguyên liệu có trong hình ảnh. Từ đó tính khoảng cách và xác định được tên món ăn. Kiểm thử hoạt động của nút này được trình bày ở hình 31;
- “Các thành phần nguyên liệu gồm:” đây là kết quả tập hợp các thành phần nguyên liệu có trong hình ảnh mà mô hình AI nhận dạng được.



Hình 30: Kết quả upload hình ảnh



Hình 31: Kết quả xác định món ăn

## PHẦN KẾT LUẬN

### **1. Kết quả đạt được**

Để giải quyết bài toán “Xây dựng hệ thống xác định món ăn đặc sản Việt Nam dựa trên các thành phần nguyên liệu”, chúng tôi đã thực hiện quá trình cài đặt và lựa chọn mô hình máy học. Ở đây, bài toán đã được giải quyết với thị giác máy tính, cụ thể là phương pháp phát hiện đối tượng. Những đóng góp chính của niêm luận này là:

**Xây dựng được tập dữ liệu các món ăn đặc sản Việt Nam.** Để có được bộ dữ liệu này, chúng tôi đã thực hiện thu thập những hình ảnh về món ăn đặc sản Việt Nam. Nguồn ảnh được tải từ Google Images, cùng với tự chụp ảnh món ăn ngoài thực tế. Bộ dữ liệu hiện tại có gần 3000 ảnh.

**Xây dựng ba mô hình phát hiện thành phần nguyên liệu món ăn gồm: Yolov8, RT-DETR, Yolov9.** Mỗi mô hình đều được cẩn thận tinh chỉnh các siêu tham số sao cho mô hình đạt được độ chính xác cao nhưng hạn chế *overfitting* (19). Kết quả huấn luyện cả ba mô hình đều có độ chính xác cao và dung lượng phù hợp, đáp ứng được yêu cầu sử dụng thực tế.

**Đề xuất được một giải thuật mới.** Chúng tôi đã nghiên cứu, thử nghiệm trên nhiều giả thuyết khác nhau, từ đó xây dựng được một công thức mới. Công thức này sẽ tính toán được “khoảng cách” của một đối tượng đến đối tượng khác, dựa trên các đặc điểm của hai đối tượng. Trong đề tài này, chúng tôi muốn xác định tên của món ăn dựa trên thành phần nguyên liệu – đây là đặc điểm của đối tượng. Chúng tôi sử dụng công thức, tính khoảng cách từ món ăn cần xác định đến từng món ăn trong bộ luật (mon ăn đã biết). Sau khi tính toán, hệ thống sẽ có được toàn bộ những giá trị khoảng cách từ món ăn cần xác định đến lần lượt các món ăn trong bộ luật, món ăn nào gần với món ăn cần xác định nhất thì tỷ lệ cao món ăn cần xác định chính là món ăn đó.

**Xây dựng được ứng dụng web để đưa mô hình vào ứng dụng thực tế.** Các mô hình AI được xây dựng, nghiên cứu, phát triển đều cần phải đáp ứng được tiêu chí ứng dụng được vào cuộc sống thực tế, hướng tới đối tượng người sử dụng. Chúng tôi đã đưa mô hình vào thực tế bằng cách triển khai AI lên website. Giao diện website cũng được thiết kế tối giản, giúp người dùng dễ dàng sử dụng hệ thống dù là lần đầu tiên tiếp xúc.

### **2. Hướng phát triển**

Bộ dữ liệu hiện tại có gần 3000 tấm ảnh và khuyến khích phát triển thêm trong tương lai. Chúng tôi dự định sẽ tiếp tục thu thập thêm dữ liệu trong tương lai để có đa dạng món ăn đặc sản hơn nữa. Đồng thời, triển khai bộ dữ liệu như một nguồn dữ liệu mở, để cộng đồng có thể tham gia vào sử dụng, đóng góp phát triển tập dữ liệu.

Đề tài này có thể xây dựng thêm chatbot hỗ trợ người dùng về công thức nấu ăn của món ăn đặc sản. Ngoài ra, còn hỗ trợ người dùng đề xuất các địa điểm du lịch nổi tiếng, đáp ứng sở thích, tính cách, mong muốn của người dùng.

Thực hiện fine tuning với nhiều mô hình phát hiện đối tượng khác như: EfficientDet (20), SSD (21), Faster R-CNN (22)... tạo nên một loạt các mô hình xác định món ăn đặc sản Việt Nam. So sánh, đánh giá các mô hình với nhau, áp dụng các mô hình vào những dự án phù hợp.

Hệ thống có thể phát triển trên nhiều nền tảng công nghệ nữa như app desktop và mobile app, hoặc lập trình nhúng vào các thiết bị điện tử. Mở rộng hơn nguồn thiết bị, phần mềm sử dụng hệ thống.

## TÀI LIỆU THAM KHẢO

1. *A Real-time Junk Food Recognition System based on Machine Learning.* **Shifat, Sirajum Munira, Takitazwar Parthib, Sabikunnahar Talukder Pyaasa, Nila Maitra Chaity, Niloy Kumar, and Md Kishor Morol.** 2021. International Conference on Bangabandhu and Digital Bangladesh.
2. *Xác định món ăn đặc sản Việt Nam với sự kết hợp của mạng học sâu và bản thẻ học.* **Mã Trường Thành, Châu Ngân Khánh, Thạch Minh Hòn, Phạm Xuân Hiền, Phan Bích Chung.** 5A, 2023, Tạp chí Khoa học Đại học Cần Thơ, Vol. 59, pp. 93-101.
3. *A review of image-based food recognition and volume estimation artificial intelligence systems.* **Konstantakopoulos, Fotios S., Eleni I. Georga, and Dimitrios I. Fotiadis.** s.l. : IEEE Reviews in Biomedical Engineering, 2023.
4. *Một số món ngon đặc sản của các tỉnh vùng Đồng bằng sông Cửu Long.* **Trần Thị Kiều Trang, Lý Thị Trà My.** 04, s.l. : Tạp chí Nghiên cứu khoa học và Phát triển kinh tế, Trường Đại học Tây Đô, 2019.
5. *Understanding of a convolutional neural network.* **Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi.** s.l. : international conference on engineering and technology (ICET), 2017.
6. *You Only Look Once: Unified, Real-Time Object Detection.* **Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi.** s.l. : Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
7. *CSPNet: A new backbone that can enhance learning capability of CNN.* **Chien-Yao Wang, Hong-Yuan Mark Liao, I-Hau Yeh, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh.** s.l. : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020.
8. *Comparison of cspdarknet53, cspresnext-50, and efficientnet-b0 backbones on yolo v4 as object detector.* **Marsa Mahasin, Irma Amelia Dewi.** s.l. : International Journal of Engineering, Science and Information Technology, 2022.
9. *YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information.* **Wang, Chien-Yao, I-Hau Yeh, and Hong-Yuan Mark Liao.** s.l. : arXiv preprint arXiv:2402.13616, 2024.
10. *An image is worth 16x16 words: Transformers for image recognition at scale.* **Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby.** s.l. : arXiv preprint arXiv:2010.11929, 2020.

11. *Detrs beat yolos on real-time object detection.* **Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, Jie Chen.** s.l. : arXiv preprint arXiv:2304.08069, 2023.
12. *A survey on performance metrics for object-detection algorithms.* **Padilla, Rafael, Sergio L. Netto, and Eduardo AB Da Silva.** s.l. : 2020 international conference on systems, signals and image processing (IWSSIP), 2020.
13. *CNN for handwritten arabic digits recognition based on LeNet-5.* **El-Sawy, Ahmed, Hazem El-Bakry, and Mohamed Loey.** s.l. : Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016 2. Springer International Publishing, 2017.
14. *Food image classification with improved MobileNet architecture and data augmentation.* **Phiphiphatphaisit, Sirawan, and Olarik Surinta.** s.l. : Proceedings of the 3rd International Conference on Information Science and Systems, 2020.
15. *Rethinking the inception architecture for computer vision.* **Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna.** s.l. : Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
16. *Very deep convolutional networks for large-scale image recognition.* **Karen Simonyan, Andrew Zisserman.** s.l. : arXiv preprint arXiv:1409.1556, 2014.
17. *Deep residual learning for image recognition.* **Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun.** s.l. : Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
18. *Densely connected convolutional networks.* **Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger.** s.l. : Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.
19. *An overview of overfitting and its solutions.* **Ying, Xue.** IOP Publishing., s.l. : Journal of physics: Conference series, 2019, Vol. 1168.
20. *Efficientdet: Scalable and efficient object detection.* **Mingxing Tan, Ruoming Pang, Quoc V. Le.** s.l. : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.
21. *Ssd: Single shot multibox detector.* **Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg.** s.l. : Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, 2016.
22. *Faster r-cnn: Towards real-time object detection with region proposal networks.* **Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun.** s.l. : Advances in neural information processing systems 28, 2015.

23. *A Review of Yolo algorithm developments.* **Jiang, Peiyuan, et al.** s.l. : Procedia computer science 199, 2022.

24. *Object detection with deep learning: A review.* **Zhao, Zhong-Qiu, et al.** s.l. : IEEE transactions on neural networks and learning systems, 2019.